RESEARCH

Validation Study of the PISA 2000, PISA 2003 and TIMSS-2003 International Studies of Pupil Attainment

Dr Graham Ruddock, Tandi Clausen-May
Chris Purple and Rob Ager
National Foundation for Educational Research

department for
education and skills
creating opportunity, releasing potential, achieving excellence

Research Report RR772

Research Report
No 772

Validation Study of the PISA 2000, PISA 2003 and TIMSS-2003 International Studies of Pupil Attainment

Dr Graham Ruddock, Tandi Clausen-May
Chris Purple and Rob Ager

National Foundation for Educational Research

ISBN 1 84478 765 6

Contents

1. Summary

In recent years England has taken part in two separate international surveys of comparative student performance involving mathematics and science. These were the TIMSS (Trends in International Mathematics and Science Study) in 1999 and 2003 and PISA (Programme for International Student Assessment) in 2000 and 2003.

Since both studies address mathematics and science, many people have given thought to whether or not the results for England should be expected to be similar in TIMSS and in PISA. Some countries have exhibited quite marked differences in performance between PISA and TIMSS, while others appear to have performed similarly. One reason for such differences might lie in the nature of the items used and their familiarity to the students taking them. The main purpose of the work reported here was to undertake a rating study of the familiarity to students in England of the items used in PISA and in TIMSS and to assess whether this differed sufficiently to be a possible cause of difference in performance.

Expert panels rated the familiarity of the concept or skill that each item assessed, the familiarity of the context in which the item was embedded, and the familiarity of the item format. These ratings related to the percentage of the target population expected to be familiar with the aspect of the item under scrutiny. The panels also gave each item a general rating for appropriateness. They did not, however, make any attempt to judge how easy or difficult students would find each item. The focus of the study was the degree to which items were familiar and appropriate, not their level of difficulty.

When the ratings were analysed, the science items used in TIMSS and PISA were found to have been given generally similar ratings for the three aspects of familiarity examined and for appropriateness. In mathematics, however, the TIMSS items were rated as more appropriate and likely to be more familiar in context and item format than the PISA ones.

For science the average ratings for familiarity suggested that 40% to 50% of students would be familiar with what was being asked of them in both TIMSS and PISA. In mathematics, average ratings for the familiarity of the TIMSS items ranged between 65% and 85 %, while for PISA the range was 50% to 70%. The average values found reflect the differentiation present in the

English national curriculum at key stages 3 and 4, with some topics familiar only to higher attaining students.

The most significant difference found between PISA and TIMSS was the amount of reading required in PISA, which was also greater than that required in the key stage 3 tests or GCSE examinations in England. Often this high reading demand was found to be coupled with a relatively low level of mathematics or science demand. This combination would be expected given PISA's focus on the assessment of mathematical and scientific literacy. The level of mathematics that students can apply in unfamiliar contexts, for example, is lower than that they can access exactly as they have met it before or in a familiar context or application. A high reading demand was not evident in TIMSS and this difference is large enough to be a possible cause for different performance between PISA and TIMSS. The ability of students to undertake the reading required in PISA, and their willingness to do so, are important factors which are likely to affect performance in PISA items.

Direct comparison between PISA, TIMSS, key stage 3 tests and GCSE revealed that all were drawing on the same bank of item types – multiple choice, short response and more extended response – but that the balance between these types differed. Students in England who have experience of key stage 3 tests and/or GCSE would therefore be familiar with each type. TIMSS makes by far the greatest use of multiple choice items, with over 50% in this format, but these are also found in the other three assessments. In all four assessments short response items find substantial use. Extended tasks are a feature of PISA, particularly in science, but are present in the other three assessments in varying degrees.

The PISA approach, assessing mathematical or scientific literacy leads directly to assessment in context. This makes the largest difference in mathematics, where TIMSS, KS3 and GCSE all present out of context items in varying degrees, but PISA does not. Students in England are, however, familiar with both in and out of context mathematics. There is no prima facie reason why PISA's use of context should either favour or disadvantage students in England on the grounds of familiarity, and the findings of the study described here do not indicate that it did so.

2. Introduction

The participation of England in recent international comparative studies has raised interest and led to questions about the nature of the tasks assigned to students and their familiarity to students in England. England has participated in the most recent TIMSS (Trends in International Mathematics and Science Studies) in 1999 and 2003 and in the PISA (Programme for International Student Assessment) studies in 2000 and 2003. The studies differ in the age groups assessed, with Year 9 in TIMSS but a combination of Years 10 and 11 in PISA. Assessment of mathematics and science is common to both surveys, but on a rather different basis. The TIMSS surveys are overtly curriculum based, while PISA emphasises literacy, using the term to apply to science and mathematics as well as reading, all three being assessed. It is not clear whether similar results should be expected from TIMSS and PISA, so this study was designed to examine this issue and to provide a briefing on what has been tested and how familiar it is to English students.

To achieve this, one aspect of the study involved an examination of the items used in recent TIMSS and PISA studies and an evaluation of their likely familiarity to English students. The study has also provided a summary of the item types used, the features of the tests employed and the nature of what was tested. Taken together these different strands of information allow the results of both PISA and TIMSS studies to be interpreted in the light of more complete knowledge of what has been assessed and how it has been tested, in order to judge whether or not similar results should be expected from the two international surveys.

The emphasis in this study was on the secondary age range. The TIMSS mathematics and science surveys are grade based and the data derives from Year 9 in England (the concurrent assessment of Year 5 in TIMSS is not considered here). PISA is age based and the data is from students in Years 10 and 11. In the TIMSS structure, half of the items assess mathematics while the other half assess science. PISA has a more complex structure, with three domains assessed in each survey: reading literacy, mathematical literacy and scientific literacy. In each survey, one of these domains is the major one, taking up the majority of the assessment. In PISA 2000 reading literacy was the focus, while in 2003 this shifted to mathematical literacy. Science will be the

focus in 2006, but since science has not yet been the major part of a PISA assessment there are fewer science items available for review.

In this report the familiarity rating exercise is presented first, with the results in section 4. Section 5 deals with other aspects of the tests found to be important when reviewing items, and a general summary of findings is given in section 5.5. Section 6 contains a direct comparison between TIMSS, PISA, key stage 3 tests and GCSE examinations in mathematics and science, providing a further insight into similarities, differences and familiarity to students in England. There is no TIMSS study comparable to the PISA study of reading literacy, so the section compares PISA, key stage 3 English and GCSE English.

The appendices show a range of released TIMSS and PISA items which supplement those shown and discussed earlier and are intended to provide the reader with a better feel for the nature of the TIMSS and PISA tests.

3. The Rating Process

NFER staff developed a rating system which would allow both TIMSS and PISA items to be rated for familiarity. The work followed on from that carried out in Ireland on PISA items but with modifications to suit the nature of the English curriculum and to meet the requirement to make comparisons between PISA items and TIMSS items[1]. Since the two surveys test students from different key stages, and hence programmes of study (curricula), a system which involved the percentage of students expected to be familiar with the aspect of the item concerned was developed.

Teams of consultants were provided by QCA to rate the familiarity of the TIMSS and PISA items. Those selected were very experienced in assessment at both key stages 3 and 4, being familiar with both key stage 3 tests and GCSE examinations in their subject area.

The teams of consultants for mathematics and science were asked to rate each part of each question in both TIMSS and PISA for 'familiarity' from four different aspects. A third team performed a similar task, but with slightly different criteria, for the PISA reading literacy questions. Further detail is given below.

All three teams were asked first to give an overall rating of the appropriateness of the question for their own students in the relevant year group, from 'Not at all appropriate' (1), to 'Very appropriate' (5).

	Do you feel that this question would be generally appropriate for your students?				
	Not at all appropriate				Very appropriate
Code	**1**	**2**	**3**	**4**	**5**

Then the mathematics and science consultants were asked to rate the familiarity of the *concept*, the *context* and the *format* of each part of each question. The reading literacy consultants were

[1] See Ready for Life? The literacy achievements of Irish 15 year-olds with comparative international data: Shiel, Cosgrove, Sofroniou and Kelly, ERC, 2001

asked to make a similar rating, but for the *skill*, *text type* and *task* rather than the *concept*, *context* and *format*. The rating scale and the codes used are shown below.

	Not in Curriculum	**Percentage of Students Estimated to be Familiar with the Concept/ Skill, Context/Text type, or Format/Task**				
	None	Very few	Fewer than half	About half	More than half	All or almost all
	0%	about 5%	about 25%	about 50%	about 75%	about 95%
Code	**0**	**1**	**2**	**3**	**4**	**5**

In addition, the mathematics consultants were asked to indicate whether the availability or otherwise of a calculator was appropriate for students in England, on a question-by-question basis.

The definitions of the factors being rated evolved during discussion with the three groups of consultants. The approaches varied somewhat to reflect the focus of the different subject areas. These are summarised below.

Definitions of 'Familiarity' – Mathematics and Science

Familiar with

- the concept:
  Students who are familiar with a mathematical or scientific concept have studied it in class, and feel confident that they understand it.

- the context:
  Students who are familiar with the context of a question understand any 'story' and /or graphics associated with it, and any mathematical/scientific context. They feel confident that they know what the question is asking.

- the format:
  Students who are familiar with the format of a question have met similar questions in the same format, presented in the same sort of language.

In addition, the mathematics consultants observed that

> *The format rating is also affected by whether students are accustomed to meeting questions where the difficulty has been significantly increased by testing diverse skills within the same question.*

The reading literacy consultants devised a more complex rating structure, as shown in the table on the following page. It is clear that in this subject area consultants found it more difficult to distinguish clearly between the *familiarity* of the texts and questions and their level of *difficulty*. This problem was not so evident from the discussions of the mathematics and science consultants, although it was still present. It is possible to set a hard or an easy computation, for example, in the same context and format, so two mathematics problems may be rated as equally familiar despite showing very different facilities. In reading literacy, on the other hand, students are normally exposed only to material that they are likely to be able to understand. This means that, as a generalisation, a demanding piece of text is likely to be familiar to fewer students than one that is more accessible, so the familiarity ratings relate more directly to the perceived difficulty of the materials.

The process was initiated at a half day meeting when all nine raters were briefed, discussed the brief, clarified it and rated trial batches of items. The clarification and querying process then continued via Email as the raters completed the rating tasks at home.

Ratings were entered directly into Excel spreadsheets. These were transmitted to NFER where the ratings for each subject were compared. NFER staff then ordered the items by the degree of disagreement between the raters. The raters then met again at a second half-day meeting and the items with high levels of disagreement were reviewed and, where appropriate, ratings were amended after discussion between the raters and the NFER team. The amended ratings were then analysed as detailed in section 4.

.

PISA reading tests: analysis of grades given. Y10/11 reading ability Reasons for grading

Aspect	Suitability	Skill	Text	Task
5	**Texts and skills feature in EN teaching at all ability levels and the task is approachable.** accessible to almost all students (95%)	**widely taught** applied successfully by almost all students	**of a type accessible to almost all** easy to follow, familiar type, few complex sentences and clusters of abstract nouns; text self sufficient or well contextualised; topic clearly signalled	**clear and unambiguous** evidence for the question is easily obtained from the text without much reinterpretation or deal with abstract terms.
4	accessible to most students but at least 25% will find it hard to do.	**widely taught but not necessarily part of mainstream English** applied successfully by most students but inaccurately or imprecisely by a significant minority	**of a type widely used in English teaching:** **OR** **not of a type widely used in English but easy to read** e.g. with clear headings and separate boxes	**clear but requiring some inference or judgements about relevance**
3	accessible to only about half of students	**widely taught but involving higher level skills** applied unreliably or with difficulty by many students	**of a type widely encountered but linguistically or conceptually quite hard** **OR** **not of a type widely used in English but fairly easy to read** e.g. tables with limited context to offset some difficulties in language	**clear but of a type that students find hard** **OR** **quite complex but answer located in easily identified discrete portion of the text** **OR** **very complex but the information is easily accessible** **OR** **requiring application of numeracy as well as literacy skills**
2	accessible to only the top 25% of students	**requiring high level skills of a sort unlikely to be taught to all students at this stage** conceptually demanding and beyond the competence of most students	**of a familiar text type but very hard to read or difficult to contextualise** **OR** **a text type sometimes used in English but this example relies heavily on general or other subject knowledge**	**demanding and requiring complex or high-level inference** e.g. the task draws more on numeracy than literacy skills but may be familiar to some students in the context of other subject disciplines
1	**Texts and skills are of a sort that few students are likely to have encountered or will find accessible** accessible to very few students (5%)	**requiring levels of experience and understanding more likely to be encountered in post-16 students**	**very obscure**	**difficult to answer using only the information provided** **OR** **the task is presented in such a way as to confuse many students**
0	**Accessible to virtually no students at this level**			

4 Results of the Familiarity Rating Process

4.1 Introduction to the Results

The results of the familiarity rating exercise are shown below in a series of tables. In each table the mean ratings for appropriateness of the item and for the familiarity of the concept being tested, the context used and the item format are given. The standard deviations are also given and represent the amount of variation in appropriateness or familiarity.

The initial tables give the ratings for all relevant items in either PISA or TIMSS, while the later ones break the items down by topic, where there are sufficient items in each topic to do so, and, for TIMSS, by whether the item was first used in 1995,1999 or 2003.

Given the coding system employed, the higher the mean value, the greater the perceived level of familiarity was. A mean of around 3 indicates that the item was viewed as being familiar to about half of the target students. Values lower than 3 correspond to fewer than 50%, with a mean of 2 representing around 25% of students. Similarly, a value of 4 represents around 75% of students. In most cases the mean ratings were between 3 and 4, representing familiarity to between 50% and 75% of students, most usually closer to 50% than to 75%.

4.2 Ratings at Subject Level

Table 4.1, below, summarises the ratings for the science items in both TIMSS and PISA.

Table 4.1 Familiarity ratings for TIMSS and PISA science items

	Mean Score					Standard Deviation				
Science	Is the Item Appropriate	Familiarity of Concept	Context	Format		Is the Item Appropriate	Familiarity of Concept	Context	Format	Number of items
TIMSS	3.36	3.07	3.07	2.81		1.38	1.55	1.46	1.45	268
PISA	3.23	3.06	3.05	2.60		0.86	0.84	0.79	0.84	45

For science in TIMSS and scientific literacy in PISA, the mean values were fairly close to 3, representing around 50% of students. In both surveys familiarity of item format received the lowest ratings, below 3. The ratings for the familiarity of concept and context were very similar for both, while for appropriateness and familiarity of item format the TIMSS science items were rated slightly higher. The higher standard deviations for the TIMSS science items should be noted, since they indicate more items with high and low ratings than in PISA while the means for both surveys did not differ greatly.

Table 4.2, below, summarises the ratings for the mathematics items in both TIMSS and PISA.

Table 4.2 Familiarity ratings for TIMSS and PISA mathematics items

	Mean Score					Standard Deviation				
Maths	Is the Item Appropriate	Familiarity of Concept	Context	Format		Is the Item Appropriate	Familiarity of Concept	Context	Format	Number of items
TIMSS	4.42	3.55	4.17	3.65		0.55	0.74	0.42	0.80	278
PISA	3.49	3.46	3.69	2.92		0.87	0.90	0.72	0.83	97

Comparing the PISA mathematical literacy and the TIMSS mathematics items, the picture is one of much higher ratings for TIMSS items for appropriateness, and for familiarity of context and item format. The mean ratings were over 4 (representing 75% of students) for both appropriateness and familiarity of context in the TIMSS items. Item format for PISA was the only mean rating below 3, less than 50% of students. The reviewers perceived the TIMSS items as being more appropriate than those in PISA, and took the view that both the context and the item format used were likely to be more familiar in TIMSS. The higher standard deviations (over 1.0) noted in the ratings of the TIMSS science items were not present for TIMSS mathematics items.

Table 4.3 gives details of the ratings of the reading literacy items in PISA.

Table 4.3 Familiarity ratings for PISA reading literacy items

Reading	Mean Score					Standard Deviation				Number of items
	Is the Item Appropriate	Familiarity of				Is the Item Appropriate	Familiarity of			
		Skill	Text type	Format			Skill	Text type	Format	
PISA	3.54	3.65	3.51	3.47		0.55	0.51	0.55	0.57	144

The ratings for appropriateness and the three aspects of familiarity, in this case skill, text type and format, were very similar, around 3.5, representing around 60% to 65% of students. The standard deviations were again similar in size and low in value, indicating less variation than in science or most aspects of mathematics. Since different rating panels were used and different subject matter was under review, comparisons between the mean ratings across subjects should be made cautiously.

4.3 Ratings for TIMSS Items by Date of Introduction

The TIMSS items reviewed could be identified as first appearing in the 1995, 1999 or 2003 surveys, and these groups of item were analysed separately. Table 4.4 gives the analysis for TIMSS science items.

Table 4.4 Familiarity ratings for TIMSS science items by year first used

TIMSS by age of item Science	Mean Score					Standard Deviation				Number of items
	Is the Item Appropriate	Familiarity of				Is the Item Appropriate	Familiarity of			
		Concept	Context	Format			Concept	Context	Format	
1995	3.63	3.47	3.42	3.25		1.41	1.60	1.54	1.46	48
1999	3.38	3.08	3.13	2.83		1.32	1.47	1.37	1.38	101
2003	3.24	2.91	2.88	2.63		1.42	1.57	1.47	1.47	119
overall	3.36	3.07	3.07	2.81		1.38	1.55	1.46	1.45	268

For science the older items generally received higher appropriateness and familiarity ratings than newer ones, a somewhat surprising pattern, and one that is difficult to explain. Table 4.5 gives a similar analysis for TIMSS mathematics items

Table 4.5 Familiarity ratings for TIMSS mathematics items by year first used

TIMSS by age of item Maths	Mean Score					Standard Deviation				Number of items
	Is the Item Appropriate	Familiarity of Concept	Context	Format		Is the Item Appropriate	Familiarity of Concept	Context	Format	
1995	4.42	3.95	4.24	3.70		0.58	0.65	0.36	0.77	48
1999	4.53	3.61	4.20	3.72		0.49	0.73	0.49	0.84	115
2003	4.31	3.33	4.12	3.57		0.59	0.70	0.38	0.76	115
overall	4.42	3.55	4.17	3.65		0.55	0.74	0.42	0.80	278

For the TIMSS mathematics items, ratings by age of item were closer, but again the tendency was for the older, 1995 items to be rated as slightly more familiar than items introduced in 2003. Again this seems rather surprising.

4.4 Ratings for TIMSS and PISA Items by Topic Assessed

Where possible, the items were also analysed by the topic allocated by the test constructors. This was possible for both mathematics and science in TIMSS and for reading literacy in PISA. There were too few PISA science items to break down in this way, while the average number of items in each PISA mathematics topic was rather low. Table 4.6 shows the analysis of the TIMSS science items by topic.

Table 4.6 Familiarity ratings for TIMSS science items by topic

TIMSS items by topic Science	Mean Score					Standard Deviation				Number of items
	Is the Item Appropriate	Familiarity of Concept	Context	Format		Is the Item Appropriate	Familiarity of Concept	Context	Format	
Chemistry	3.91	3.96	3.72	3.53		1.14	1.31	1.23	1.30	34
Earth Science	2.83	2.58	2.58	2.28		1.56	1.82	1.79	1.70	42
Environmental Science	1.99	1.55	1.69	1.44		1.35	1.65	1.50	1.45	33
Life Science	3.44	3.15	3.17	2.91		1.22	1.24	1.14	1.21	81
Physics	1.10	3.52	3.54	3.29		1.10	1.23	1.23	1.17	60

When the TIMSS science items were broken down into the content areas allocated by TIMSS, distinct differences between topic areas were apparent. For environmental science items, all familiarity ratings were below 2, less than 25% of students, while earth science items were rated below 3, less than 50% of students. These lower values should, however, be treated with caution since the some of the content concerned is not in the national curriculum for science but may be covered in other subjects, such as geography, and may, therefore, have been more difficult to rate. The appropriateness rating for physics items was lower than for other areas, but this did not extend to the familiarity ratings. Standard deviations were relatively high for all topic areas. Table 4.7 gives similar details for TIMSS mathematics items.

Table 4.7 Familiarity ratings for TIMSS mathematics items by topic

TIMSS items by topic Maths	Mean Score					Standard Deviation				Number of items
	Is the Item Appropriate	Familiarity of Concept	Context	Format		Is the Item Appropriate	Familiarity of Concept	Context	Format	
Algebra	4.59	3.10	4.02	3.54		0.41	0.73	0.47	0.57	63
Data Rep	4.23	3.72	4.07	3.35		0.57	0.63	0.53	0.64	34
Number	4.47	3.67	4.25	3.71		0.50	0.72	0.39	0.86	91
Geometry	4.41	3.48	4.29	3.65		0.58	0.67	0.35	1.01	39
Measurement	4.44	3.86	4.24	4.09		0.50	0.68	0.35	0.72	39

The TIMSS mathematics items, on the other hand, did not show much variation in the mean ratings when analysed by topic. Overall, neither the perceived appropriateness of the items nor their estimated familiarity seemed to be strongly associated with topic. Table 4.8 gives similar information for PISA reading literacy items.

Table 4.8 Familiarity ratings for PISA reading literacy items by topic

PISA items by topic Reading Literacy	Mean Score					Standard Deviation				Number of items
	Is the Item Appropriate	Familiarity of Skill	Text type	Format		Is the Item Appropriate	Familiarity of Skill	Text type	Format	
Interpreting	3.61	3.65	3.51	3.52		0.51	0.46	0.56	0.54	73
Reflecting	3.43	3.39	3.48	3.20		0.72	0.71	0.63	0.72	29
Retrieving info.	3.52	3.81	3.52	3.56		0.44	0.37	0.47	0.41	42

When the PISA reading literacy items were analysed by the topics used by the PISA consortium, no distinctive patterns of differences were apparent. In this respect they were similar to the TIMSS mathematics items.

4.5 Summary of Findings from the Familiarity Rating Exercise

The ratings exercise allowed direct comparisons between the perceived appropriateness and familiarity of the TIMSS and PISA items addressing mathematics and science to be made. In science the average ratings were similar for both surveys, providing no evidence that differential familiarity or appropriateness among the items might be expected to lead to different levels of performance in PISA and TIMSS assessments on these grounds. In mathematics, however, the TIMSS items were rated more highly for appropriateness and for the likely familiarity level with the context of the items and their format. No such comparison was possible in reading literacy since this is not assessed in TIMSS, only in PISA.

Typically the familiarity ratings indicated that the raters expected the aspects of the items under scrutiny would be familiar to around 50% of the target students. This is not a particularly high level but, in part, reflects the differentiation present in the English national curriculum as taught to these age groups.

In reviewing and rating the TIMSS and PISA items other factors which may well impact on expected levels of performance in England were revealed. These and emerging issues discussed between the raters and the NFER team are presented in section 5.

5. Other Aspects of the TIMSS and PISA Tests

During the study the work of the consultants, the discussions between the NFER team and the consultants and the related work carried out by the NFER team led to a considerable amount of time being spent inspecting and evaluating items from both surveys. The results of these discussions and reviews, which are additional to the ratings already presented, are summarised in section 5.

5.1 PISA and TIMSS – Some similarities and some differences

Although PISA and TIMSS have a lot of similarities, they also have some significant differences. As Elois Scott, of the US National Centre for Educational Statistics, explains, 'each was designed to serve a different purpose and each is based on a separate and unique framework and set of items' (*Comparing NAEP, TIMSS, and PISA in Mathematics and Science*, http://nces.ed.gov/timss/pdf/naep_timss_pisa_comp.pdf).

A number of general differences between the PISA and TIMSS tests began to emerge during the familiarity rating exercise. It became apparent that larger scale differences between the two surveys needed to be identified and documented. These were not covered by the item reviewing process, but impacted on it.

One important difference that was documented is the scale of the tasks involved. The TIMSS surveys contain mostly short items, with a maximum of two score points (marks) awarded. A particular context, in either mathematics or science, thus normally yields only one or two score points. From 2003 TIMSS also included Problem Solving and Inquiry tasks (PSIs) involving larger contexts generating a number of score points. PISA makes much greater use of large scale contexts generating a number of score points, particularly in reading and scientific literacy. The two surveys thus differ in this important respect, and this impacts on how students are likely to view and respond to the items they are presented with.

To continue to bring out other facets of the two surveys, this section of the report focuses on just a few individual mathematics and science questions in each test, and offers a qualitative analysis of the views of the consultants and of the NFER team and of what can de drawn out from these. The items presented or discussed in this section were raised by the consultants to illustrate

particular points. Only released questions are discussed in detail, although, since the consultants' reports were based on the complete set from each test, some other points are raised in more general terms.

5.2 The Amount of Reading Involved in PISA

The PISA tests seek to assess students' problem-solving abilities within the context of mathematics, science and reading literacy. As the Introduction to OECD/PISA 2003 Assessment explains,

'while OECD/PISA does assess students' knowledge, it also examines their ability to reflect, and to apply their knowledge and experience to real-world issues.'
(The PISA 2003 Assessment Framework, pg 20)

Problem solving may be regarded as relating primarily to the solution of non-routine, unfamiliar tasks – and so, to some extent at least, the questions in PISA might validly be designed to be 'unfamiliar' to the students. However, the most unfamiliar aspect of the PISA questions that was mentioned repeatedly by the commentators was not their problem solving focus, but rather, the amount of reading that the students had to get through before they came to the mathematics or the science. There was concern that this large amount of reading might prove to be unfruitful or, in odd cases, counterproductive. For example, one scientific literacy question, *Ozone* (PISA, S253) starts:

OZONE

The atmosphere is an ocean of air and a precious natural resource for sustaining life on Earth. Unfortunately, human activities based on national/personal interests are causing harm to this common resource, notably by depleting the fragile ozone layer, which acts as a protective shield for life on Earth.

Ozone molecules consist of three oxygen atoms, as opposed to oxygen molecules which consist of two oxygen atoms. Ozone molecules are exceedingly rare: fewer than ten in every million molecules of air. However, for nearly a billion years, their presence in the atmosphere has played a vital role in safeguarding life on Earth. Depending on where it is located, ozone can

either protect or harm life on Earth. The ozone in the troposphere (up to 10 kilometres above the Earth's surface) is 'bad' ozone which can damage lung tissues and plants. But about 90 percent of ozone found in the stratosphere (between 10 and 40 kilometres above the Earth's surface) is 'good' ozone which plays a beneficial role by absorbing dangerous ultraviolet (UV-B) radiation from the Sun.'

The passage, regarded as grammatically complex by the consultants, continues with another similar paragraph. Then the first question states:

In the text above nothing is mentioned about the way ozone is formed in the atmosphere. It goes on to give yet more information on this topic, and then to ask a question about it. The science commentators suggested that some English 15-year-olds would be likely to feel that they had been 'tricked' into struggling through a mass of inaccessible and unhelpful material, only to find that they did not need it in order to answer the question. As one consultant put it,

'A lower-achieving student will look at all this reading, say something rude, and turn over.' Another added the concern that:

'Some students will ignore all the information, and try to answer the question anyhow.'

The consultants pointed out that in some cases this approach could be effective. For example, see *Daylight* (PISA, S129).

DAYLIGHT

Read the following information and answer the questions that follow.

Today, as the Northern Hemisphere celebrates its longest day, Australians will experience their shortest. *In Melbourne*, Australia, the sun will rise at 7:36 am and set at 5:08 pm, giving nine hours and 32 minutes of daylight.* *Compare today to the year's longest day in the Southern Hemisphere, expected on 22 December, when the sun will*	*rise at 5:55 am and set at 8:42 pm, giving 14 hours and 47 minutes of daylight.* *The President of the Astronomical Society, Mr Perry Vlahos, said the existence of changing seasons in the Northern and Southern Hemispheres was linked to the Earth's 23-degree tilt.*

**Melbourne is a city in Australia at a latitude of about 38 degrees south of the equator.*

Which statement explains why daylight and darkness occur on earth?

A. The earth rotates on its axis.

B. The sun rotates on its axis.

C. The earth's axis is tilted.

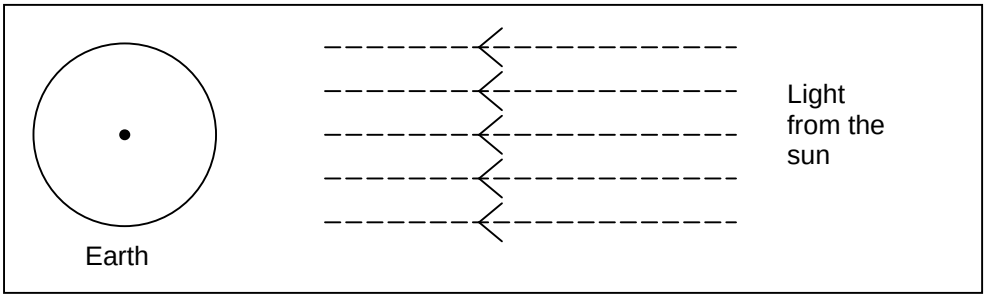
D. The earth revolves around the sun.

Reviewers were concerned that the text preceding the first question gave no information about the causes of daylight and darkness. Rather, the statement in the last paragraph appertaining to changing seasons could serve as a miscue, since it could be taken to imply that the correct response was option C, rather than option A. Students who did try to read the text might be

misled: those who ignored it might be more likely to guess correctly when they answered the question.

Similarly, the diagram in the second part of this question was regarded as having little to do with the introductory text, and was judged by the consultants to be of poor quality and difficult to interpret:

In the Figure light rays from the sun are shown shining on the Earth.



Suppose it is the shortest day in Melbourne.

Show the earth's axis, the Northern Hemisphere, the Southern Hemisphere and the Equator on the figure. Label all parts of your answer.

The science commentators themselves had some difficulty understanding what was required in this question.

Another science question which drew adverse comments, *South Rainea* (PISA S252 ***Not Released***), has not been released as it has been held back for use in future cycles of PISA. It presented a wealth of data referring to a place, apparently imaginary, called 'South Rainea', and it asked about the weather, including the amount of rain, in north and south South Rainea. In English, at least, this question about the rainier parts of north and south South Rainea was felt to present a linguistic hurdle which had little to do with students' understanding of science. It is not clear whether the same confusion between 'rain' and 'Rainea' would be likely to occur in other languages.

All the PISA science and mathematics questions were set in a context. The consultants suggested that some of the contexts used would be unfamiliar to many students in England, although not to all. For example, students would be unlikely to study the aspects of the ozone layer that are discussed in the question referred to above, or the weather and tides that are the focus of an unreleased question, within the context of the science curriculum. Some students would meet these topics in geography, but, unlike science, geography is optional at key stage 4, and not all students take it. Thus the familiarity of the context of these questions would vary, but not in a way that related to the students' achievements in science.

However, as one science consultant remarked,

'It's not the familiarity of the format, but the complexity of the language that makes PISA hard.'

The mathematics consultants expressed similar views. Unfortunately, most of the PISA mathematics questions on which they commented have not been released, but one example that may be given here is in *Exchange Rate* (PISA M413). This question first required students to use a given exchange rate to convert 3000 Singapore dollars to South African Rand. Then it continued:

On returning to Singapore after 3 months, Mei-Ling had 3900 ZAR left. She changed this back to Singapore dollars, noting that the exchange rate had changed…

The context of this question is very familiar, but as one of the mathematics consultants remarked,

'The question could have been so much simpler'.

The use of such complex phrases as 'On returning to Singapore' and 'noting that', presents a linguistic hurdle that has little to do with the mathematics. One consultant summed this up,

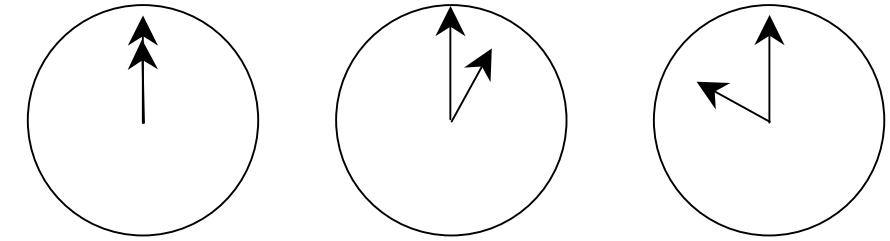
'It is the density of the questions, rather than their mathematical content, that is the hurdle. There is just too much to get into.'

In *Internet Relay Chat* (PISA M402) the context of an internet chat room is used to ask questions relating to the times in Sydney and Berlin.

INTERNET RELAY CHAT

Mark (from Sydney, Australia) and Hans (from Berlin, Germany) often communicate with each other using "chat" on the Internet. They have to log on to the Internet at the same time to be able to chat.

To find a suitable time to chat, Mark looked up a chart of world times and found the following:



Here it is both the amount and complexity of the reading demanded, and the unfamiliarity of the clock faces (which do not have any numbers), that was thought to make a fairly common-place type of question more difficult. As with a number of other questions, the consultants felt that

'In PISA, you are using maths for only a small part of what you do. Most of it is reading comprehension.'

Other questions that gave rise to a number of comments relating to the language used in the 'stories' surrounding the mathematics have not been released, so they are not given here in detail. But as one consultant remarked of the unreleased question *Number Check* (PISA M603),

'*If you take it right down to the maths they are doing, it is very, very simple. But are they ever going to get to it? It's a test of language comprehension. All they have to do is to add and subtract some numbers.'*

In general terms the amount of reading required in PISA was seen to alter the nature of the assessment, placing stress on the ability and, indeed, willingness to read large amounts of text. In comparison the level of the science or mathematics required was thought to be low.

5.3 TIMSS – Multiple Choice Items, the Number of Different Contexts and Other Issues

One aspect of the TIMSS questions that drew a lot of comment was their multiple choice structure. Although a number of the PISA items were multiple choice, the TIMSS tests made more extensive use of this type of question. Both the mathematics and the science consultants noted this, and suggested that as presented in TIMSS these were likely to be unfamiliar to some students, but not to those who had taken the Cognitive Ability Tests or assessments with a similar structure for multiple choice items, including lettered options. It was the precise format employed by TIMSS that raised discussion rather than the use of multiple choice items per se.

The consultants suggested that students who were experienced with multiple choice questions would be likely to adopt the strategy of eliminating incorrect answers in order to arrive at the correct one – a technique which works well on a number of TIMSS questions.

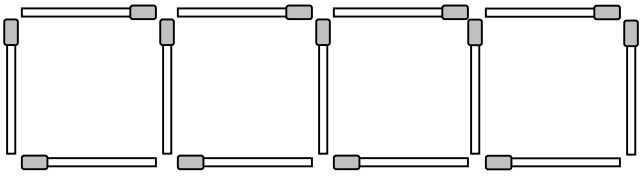
On the other hand, students who were accustomed to open-response questions, and to being asked to explain their reasoning, might be puzzled by questions to which the answer could not be found without looking through the options. Those who did not realise that there is only one correct answer to a multiple-choice question would waste time checking the others, even when they had found the correct one.

Although there were a few abstract mathematics questions in TIMSS, many of the mathematics and science questions were in context. There were fewer comments relating to the unfamiliarity of the contexts in TIMSS than in PISA, but the sheer number of different contexts was seen as a problem. In the KS3 tests, for example, a context is introduced and then, usually, three or four questions are based on it. In PISA most science and some mathematics questions followed this model, but most TIMSS questions were much shorter, with a new context introduced for each one. As one consultant remarked,

'This makes the total TIMSS experience very 'bitty' for students. They are used to a more coherent set of questions. PISA tests are different to the tests that the students know, but the questions are more familiar from the teaching programme. The PISA language is harder than students are used to, but at least the shorter questions in PISA are more familiar than the TIMSS

questions. The process of starting with a short piece of text or information, and answering a number of questions from it, as in PISA, is more familiar than the constant changes of context in TIMSS.'

For example, TIMSS MO32640 was thought to be very familiar to Y9 students, but it would normally form part of a longer series of questions developing out of the given context, and probably starting with easier lead-in questions too act as scaffolding.



In the figure, 13 matches were used to make 4 squares in a row. What is the number of squares in a row that can be made in this way using 73 matches? Show the calculations that lead to your answer.

Similarly, in question TIMSS MO22251 students are taken straight in, with no scaffolding.

Graham has twice as many books as Bob.
Chan has six more books than Bob.
Bob has x books.
Which of the following the total number of books the three boys have?

A. 3x + 6; B. 3x + 8; C. 4x + 6; D. 5x + 6; E. 8x + 2

The consultants felt that this question was valid, but that the lack of scaffolding raised its level of demand.

Both the science and the mathematics consultants commented on aspects of the language and the presentation of a number of specific questions from the TIMSS tests. One mathematics consultant remarked,

'The presentation is often more demanding than the mathematical content of the question.' For example, in TIMSS MO12016 the wording and layout were thought to complicate what was otherwise a quite simple piece of mathematics.

In which of these pairs of numbers is 2.25 larger than the first number but smaller than the second number?

A. 1 and 2; B. 2 and $^{5}/_{2}$; C. $^{5}/_{2}$ and $^{11}/_{4}$; D. $^{11}/_{4}$ and 3

This question required students to mentally place 2.25 in position between each pair of numbers in the options, in order to compare their values. The same mathematics could be assessed more simply by asking,

Which of these sets of numbers is in order?

A. 1, 2.25, 2; B. 2, 2.25, $^{5}/_{2}$; C. $^{5}/_{2}$, 2.25, $^{11}/_{4}$; D. $^{11}/_{4}$, 2.25, 3

Again, the mathematics consultants felt that the language in an algebra question, TIMSS MO22166, made it difficult to decipher.

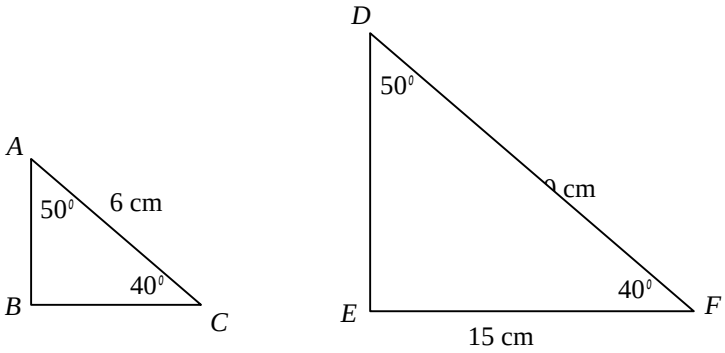
If k represents a negative number, which of these is a positive number?

A. k^2; B. k^3; C. 2k; D. $^{k}/_{2}$

The conditional 'if' was a complexity that students would not expect in a mathematics question. The change from 'k *represents* a negative number' to the expression k^2 which '*is* a positive number' was also confusing. The question itself involved a straightforward and reasonably familiar piece of mathematics, but its presentation would make it obscure for some students.

The mathematics consultants felt that the formal language used in TIMSS MO22160 would present another hurdle to some students.

The figures represent two similar triangles. The triangles are not drawn to scale.



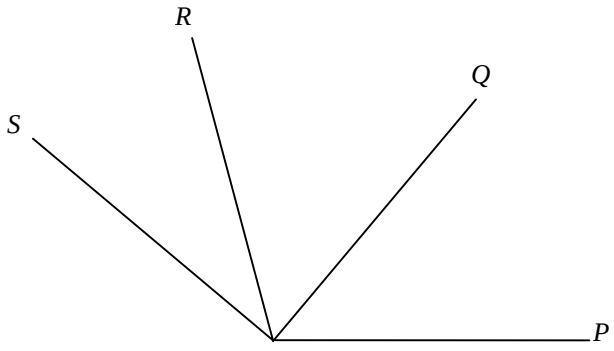
In the actual triangle ABC, what is the length of side BC?

A. 3.5 cm; B. 4.5cm; C. 5 cm; D. 5.5 cm; E. 8 cm

In a Key Stage 3 test the question would refer to 'Triangle A' or 'the shaded triangle', not 'triangle ABC'.

Similarly, TIMSS MO22202 gives a diagram, then lists a lot of information that could have been put onto the diagram but was not.

In the figure, angle POR is 110º, angle QOS is 90º, and angle POs is 140º.



What is the size of angle QOR?

Answer: _____

The student's first task is to transfer all the information onto the diagram. This was seen as a hurdle for English students, who would not be familiar with the conventions used here to refer to the angles. Once the information was available, the actual mathematical task was quite straightforward. As one consultant remarked, *'It's a complicated format to test quite simple concepts'*.

Another issue was raised by a science question, TIMSS S022152:

What processes take place in the human body that prevent it from overheating during exercise?

The science consultants queried whether students would understand that 'sweating' is being referred to here as a 'process'. Any response mentioning sweating would be awarded the mark in this question, but many students would be likely to think that a much more detailed response, describing the whole 'process' by which the sweat glands are triggered to secrete sweat, might be required.

The science consultants also queried the comparability of demand of this question in different languages. The same point was raised in relation to TIMSS SO22157:

Animals hibernate to survive cold weather and poor food supplies. Which of the following occurs in animals when they hibernate?

A. Their blood stops circulating

B. Their body temperature increases.

C. Their body fat remains constant.

D. Their rate of metabolism decreases.

The key words in this question are 'hibernate' and 'metabolism'. The consultants were not clear whether these words exist in all the other test languages, and whether they are of comparative difficulty. English has a lot of technical terms, which are scientifically accurate but relatively inaccessible. Students working in languages blessed with fewer 'scientific' terms could be at an

advantage in answering any question where the student's knowledge of technical vocabulary is crucial to their understanding of the question.

5.4 Issues Relating Specifically to Mathematics

While most of the issues that were raised by the mathematics consultants were also mentioned by the science consultants, and vice versa, there were a few points that arose in the discussion of the mathematics questions that had not been mentioned in the earlier meeting to discuss the science questions.

The mathematics consultants observed that in the TIMSS questions the amount of space given for working did not reflect the amount needed, but the science consultants did not raise this issue.

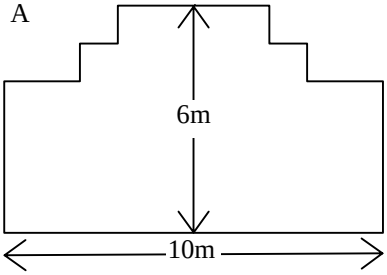
Again, the mathematics consultants were asked to comment on the availability of a calculator for all the PISA and for certain TIMSS questions, and its absence in other TIMSS questions. As expected, they identified some TIMSS questions in which a calculator would normally have been made available in England. However, there were also a few questions in each test where its availability was seen as possibly problematic. Students would not normally have access to a calculator when answering a question involving a computation with fractions, since some calculators do support such computations, but many do not, so their use would not be equitable. Thus the availability of a calculator for such questions would be unfamiliar to students in England.

Again, a calculator was available for the PISA question *Carpenter* (PISA M266).

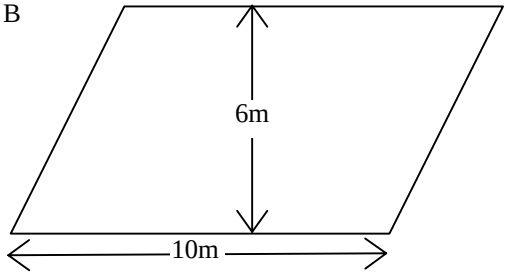
CARPENTER

A carpenter has 32 metres of timber and wants to make a border around a flower bed. He is considering the following designs for the flower bed.

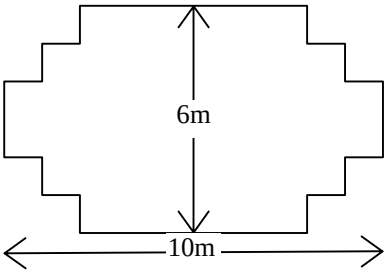
A



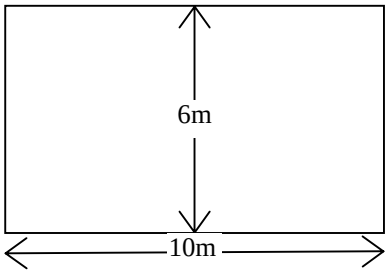
B



C



D



Circle either "Yes" or "No" for each design to indicate whether the flower bed can be made with 32 metres of timber.

Flower bed design	***Using this design, can the flower bed be made with 32 metres of timber?***
Design A	*Yes/No*
Design B	*Yes/No*
Design C	*Yes/No*
Design D	*Yes/No*

This question is best approached through spatial reasoning – the perimeter of shapes A and C may be seen to be the same as that of their surrounding rectangles. However, it was felt that giving the students access to a calculator could be a miscue to use Pythagoras. Furthermore, students who have strong spatial ability but find it difficult to interpret complex language would be likely to struggle to interpret even the quite small quantity of text surrounding this question, with terms such as 'considering', 'following designs', and 'indicate'. Finally, the labels 'A.', 'B.', 'C.', and 'D.' are similar to those used for the options in a multiple choice question, which could cause confusion.

5.5 PISA Reading Literacy

Since the reading literacy items were present only in PISA, the nature of the rating exercise was rather different. Comparisons between surveys were not made, but the comments offered by the consultants often contained echoes of points made about mathematics or science. One reason for this was that the questions went beyond what would be expected to appear in a UK reading literacy test. The consultants view was that *'a high level of numeracy was demanded'*, with *'a significant requirement to interpret diagrams'*, and that the use of context was such that some items could appear in a mathematics or science test. The range of text types was seen to be familiar, but with the caveat that some would normally be seen in subject areas other than English.

A further issue already encountered was frequent change of context / stimulus material. This had been a comment about TIMSS mathematics, rather than PISA, but the consultants view was that in reading literacy the students were faced with more texts, but fewer questions about each text, than in, say, a key stage 3 English test or other tests likely to have been taken. This was described as *'an unfamiliar format, with one thing after another'* and students having to *'jump between text types'*. Both the changes of context and the wide definition of reading literacy were seen as bringing an unfamiliar feel to students from England.

A further echo of the comments made by science and mathematics consultants was that the tests required great *'reading stamina'*. Their view was there was a lot to read, a conclusion influenced by the lower number of questions per text than in England and the consequent higher number of texts to be read. There were also concerns about difficult words which had to be read but did not

need to be understood to answer the questions – *'if they can tolerate not understanding the technical words, they can answer the questions'*.

A further point about familiarity was the authenticity of the materials. Leaflets were shown in the page of the test booklet rather than being printed separately and colour printing was not employed.

5.6 PISA and TIMSS – Common issues

Both the TIMSS and the PISA tests contained a mixture of mathematics and science questions. In addition, PISA had a number of questions assessing 'reading literacy'. Students in England would not be used to finding a mixture of science, mathematics and reading comprehension questions in the same test, but this was not thought by any of the commentators to be likely to cause significant problems.

The type-face used in both TIMSS and PISA is much smaller and less clear than that used in Key Stage 3 or GCSE papers. The TIMSS questions even use a serif font, although the PISA font is sans-serif. This in itself was felt to present a hurdle to lower-achieving students, who would be much less likely to persevere with complex sentence structures presented in inaccessible print.

In a key stage 3 test or a GCSE exam, students are told how many marks are available for an answer, and would be expected to gauge from this how much detail was expected, for example in an explanation. However, both TIMSS and PISA contained questions in which explanations were required, but no indication was given of the level of detail needed for the award of the marks. Both the science and the mathematics consultants felt that English students would be likely to spend a lot of time on these questions, and to give more elaborate responses than were required.

Generally, the consultants saw many of the problems in both TIMSS and PISA as relating to the language. The question is bound to be raised, to what extent these problems would arise in other languages where the technical terms of TIMSS, and the elaborate sentence structures of PISA, might have to be significantly simplified.

5.7 A Summary - The PISA and TIMSS Tests Compared

In order to provide an overview of the findings of the study the tables which follow summarise the structure and content of the PISA and TIMSS tests. Table 5.1 summarizes the structures, while table 5.2 summarizes the findings of the study.

Table 5.1 The structure of the PISA and TIMSS tests

PISA	TIMSS
Survey Cycle Every 3 years, 2000 onwards.	Every 4 years, 1995 onwards.
Students Aged 15 – mixture of Years 10 and 11.	In Year 9.
Papers Each student takes one paper and pencil test, total 2 hours, with a break after one hour.	Each student takes one paper and pencil test, 1½ hours, split into two 45 minute sections.
Item types A mixture of multiple choice and constructed response items.	A mixture of multiple choice and constructed response items.
Contextualisation 'The items are organised in groups based on a passage setting out a real-life situation.' OECD: The PISA 2003 Assessment Framework, p11 Most PISA questions, whether they focus on maths, science, reading or problem solving, are introduced with an extensive passage to be read.	In most cases a new context is introduced for each question, although there are a few longer problem-solving tasks with several related parts. Contexts are introduced briefly at the beginning of each question.
Domains of the tests Mathematical, Scientific and Reading Literacy Each paper contains a mixture of questions from different domains. One domain is the major one in each survey, comprising the majority of the items.	Mathematics and Science Each paper contains a mixture of questions from different domains. Each survey comprises 50% mathematics and 50 % science.
Curriculum coverage 'OECD/PISA is not constrained by the common denominator of what has been specifically taught in the schools of participating countries.' OECD: The PISA 2003 Assessment Framework, p14	'A detailed questionnaire to participating countries about topics included in their curricula provided valuable feedback on the suitability of assessing individual mathematics and science topics.' TIMSS Assessment Frameworks and Specifications 2003, p5

PISA	TIMSS
Pedagogical approach 'OECD/PISA does not exclude curriculum-based knowledge and understanding, but it tests for it mainly in terms of the acquisition of broad concepts and skills that allow knowledge to be applied.' OECD: The PISA 2003 Assessment Framework, p14	'From 2003 on, TIMSS will gradually place more emphasis on questions and tasks that offer better insight into students' analytical, problem-solving and enquiry skills and capabilities.' TIMSS Assessment Frameworks and Specifications 2003, p5
Calculators in mathematics A calculator is available throughout the test.	The first half of the test is non-calculator; a calculator is available for the second half.

Table 5.2 The familiarity and other important features of the PISA and TIMSS mathematics and science assessments compared

PISA	TIMSS
Science Items PISA and TIMSS items seen as very similar in appropriateness and familiarity. In both, items rated as appropriate for over 50% of students, around 50% of students estimated to be familiar with the concept assessed and the item context. Item format estimated to be familiar to over 40% of students.	
Mathematics Items TIMSS items seen as more appropriate and familiar in context and format than PISA items. The concept assessed in both PISA and TIMSS estimated to be familiar to around 60% of students.	
PISA mathematics items Seen as appropriate for about 60% of students Context familiar to nearly 70% of students Item format familiar to nearly 50% of students	*TIMSS mathematics items* Seen as appropriate for about 85% of students Context familiar to about 80% of students Item format familiar to about 65% of students
PISA	**TIMSS**
Reading Demands A lot of text to read, much more than in English practice; grammatical complexity also an issue.	***Reading Demands*** Amount to be read usually similar to English practice, some awkward or complex phrasing.
Subject Demands The mathematics and science required often relatively low level as it is often being applied in a context new to the student.	***Subject Demands*** The mathematics and science required often similar to practice in England.
Use of Context Usually large scale contexts with a number of marks related to them, but moderate number of contexts for students to deal with.	***Use of Context*** Usually small scale contexts with one or two marks related to them, and, compared with key stage 3 tests, a large number of contexts for students to deal with.

Given the differences between the two surveys identified above, and particularly because of the high reading demands of PISA and the often low level of mathematics or science required in the PISA contextualised approach, it might be expected that the performance of some countries would differ between PISA and TIMSS. In some countries, for example, students might find the amount of reading required in PISA to be a disadvantage, while in others this might not be the case while the relatively lower level of mathematics or science often required might be advantageous.

We understand that for the 2006 assessment the PISA consortium are trying to reduce the reading requirements. If this is achieved, the likelihood of differential performance between PISA and TIMSS for other than curricular reasons might be reduced.

6 TIMSS, PISA Key Stage 3 Tests and GCSE Examinations Compared

The likely familiarity of TIMSS and PISA items as individual items has already been established and discussed and the purpose of this section is to compare these two surveys with the tests and examinations with which students in key stages 3 and 4 are familiar. The rating exercise looked at each item separately, while here the emphasis is on the general feel of the test or examination paper, including the absence or presence of a context for the mathematics being assessed, and the number of marks derived from given contexts or scenarios.

6.1 Question Styles in the Mathematics Tests

Overall, there is not a clear-cut 'PISA-style', 'TIMSS-style', 'GCSE-style' or 'KS3-style' of mathematics question. Each cycle of both PISA and TIMSS contains a wide range of questions of different types, and in particular, lengths. These overlap to a great extent with the range of GCSE and KS3 maths questions. The proportions of different types of question are different in the various assessments, but the overall ranges have many similarities.

6.1.1 PISA Mathematics Questions

PISA mathematics covers a range of question types and lengths. The longest fill a double page spread, many cover 1½ pages but there are also 1 page, ½ page and short questions. All the PISA questions are in context, as would be expected given the intention to assess mathematical literacy.

Some, but not all, PISA mathematics questions are fairly complex and carry a number of marks. For example, each cycle of the PISA tests included a question that required pupils to identify the structure of a sequence of patterns, and to interpret this numerically or algebraically. In 2000 the question on this topic, *Apples* (see below), carried 5 marks, with two related sequences of patterns. Pupils are first required simply to continue the two patterns of numbers.

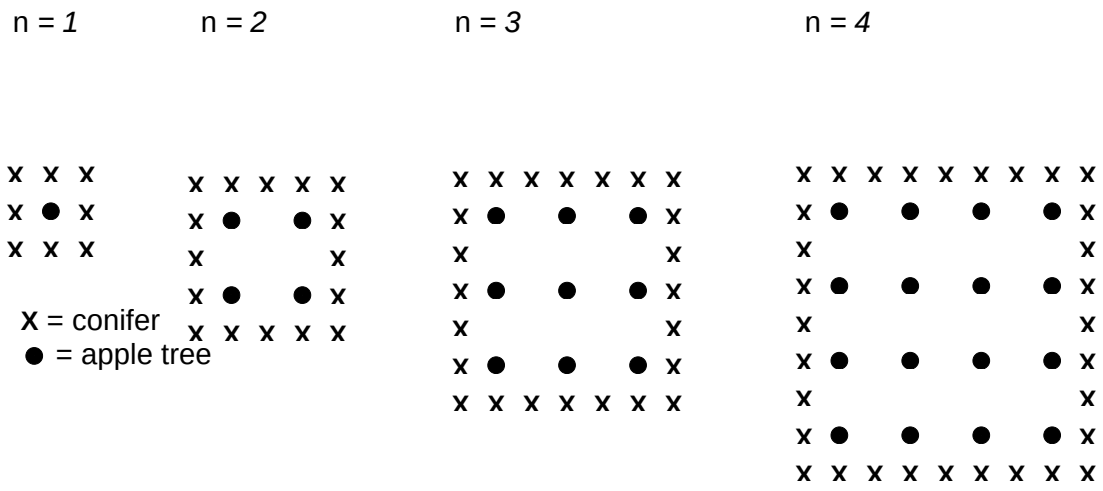
In the second part of the question pupils are asked to find n such that the values of n^{th} member of the two sequences of patterns are equal. The rules for the sequences are defined algebraically, and pupils are prompted to take an algebraic approach to the task although other methods based on geometric reasoning or on numerical trial and error are also accepted.

Finally, pupils are asked to compare the two formulae, for the sequence of patterns of apple trees, and of conifers.

PISA, Apples, 2000 only

A farmer plants apple trees in a square pattern. In order to protect the trees against the wind he plants conifers all around the orchard.

Here you see a diagram of this situation where you can see the pattern of apple trees and conifers for any number (n) of rows of apple trees:



Complete the table:

n	Number of apple trees	Number of conifers
1	1	8
2	4	
3		
4		
5		

There are two formulae you can use to calculate the number of apple trees and the number of conifers for the pattern described above:

Number of apple trees = n^2

Number of conifers = $8n$

where *n* is the number of rows of apple trees.

There is a value of *n* for which the number of apple trees equals the number of conifers. Find the value of *n* and show your method of calculating this.

Suppose the farmer wants to make a much larger orchard with many rows of trees. As the farmer makes the orchard bigger, which will increase more quickly: the number of apple trees or the number of conifers? Explain how you found your answer.

.....

.....

Apples offers an example of a complex, multi-part PISA question carrying a total of five marks. On the other hand, this may be compared with *Step Pattern*, a released question on the same topic from the 2003 cycle of PISA which follows. This is much simpler than the 2000 question.

PISA, Step Pattern, 2003 only

Question 1: STEP PATTERN M806Q01

Robert builds a step pattern using squares. Here are the stages he follows.

Stage 1 Stage 2 Stage 3

As you can see, he uses one square for Stage 1, three squares for Stage 2 and six for Stage 3.

How many squares should he use for the fourth stage?

Answer: ..... squares.

Step Pattern is equivalent to only the first step in the first part of *Apples*, and it carries only one mark.

This pair of questions indicates the extent to which PISA questions vary. Both are open, requiring some degree of constructed response to be made. However, the 2003 question was shorter and simpler than the 2000 question on this particular topic.

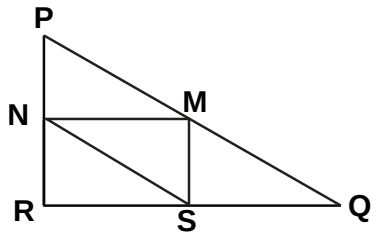
PISA also has a number of multiple choice questions. *Triangles*, from the 2000 cycle, offers an example.

PISA, Triangles, 2000 only q M161

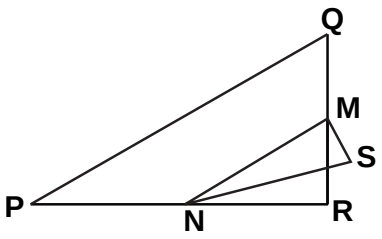
Circle the one figure below that fits the following description.

Triangle PQR is a right triangle with right angle at R. The line RQ is shorter than the line PR. M is the midpoint of the line PQ and N is the midpoint of the line QR. S is a point inside the triangle. The line MN is longer than the line MS.

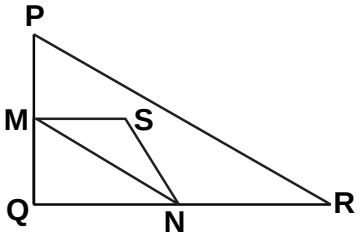
A



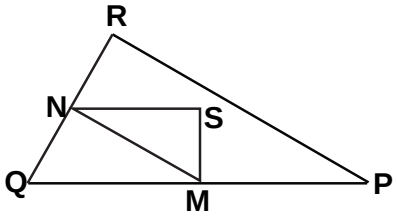
B



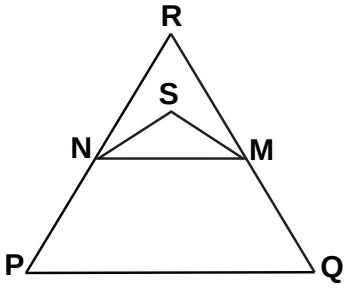
C



D



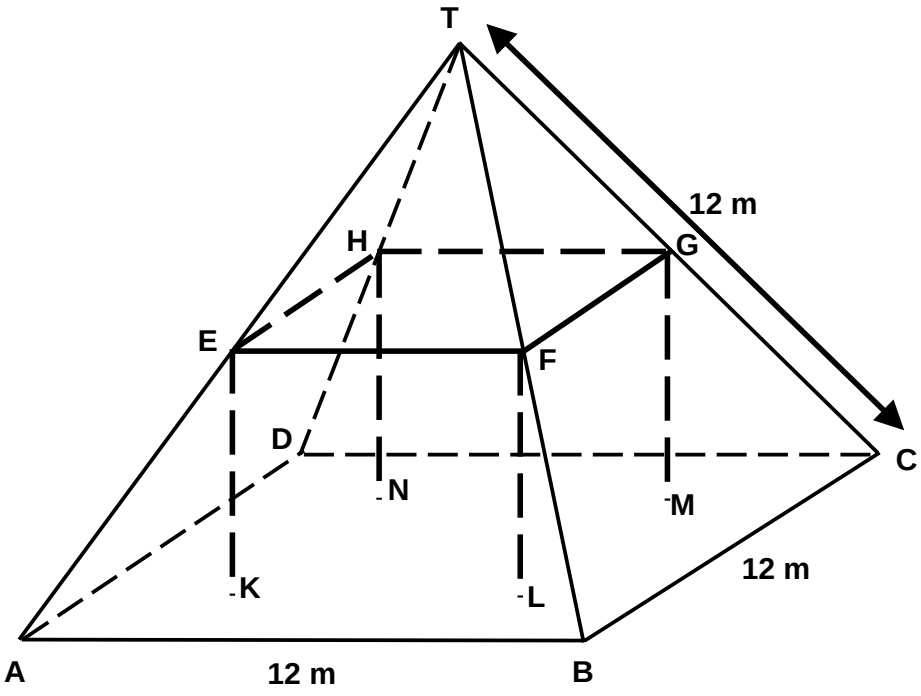
E



Another aspect of some (though not all) PISA mathematics questions is the inclusion of a great deal of extraneous information which the pupil must recognise as irrelevant and then ignore. A released question of this type is *Farms*, from the 2000 PISA cycle.

PISA, Farms, 2000 only q M037

Here you see a photograph of a farmhouse with a roof in the shape of a pyramid. Below is a student's mathematical model of the farmhouse **roof** with measurements added.



The attic floor, ABCD in the model, is a square. The beams that support the roof are the edges of a block (rectangular prism) EFGHKLMN. E is the middle of AT, F is the middle of BT, G is the middle of CT and H is the middle of DT. All the edges of the pyramid in the model have length 12 m.

Calculate the area of the attic floor ABCD.

The area of the attic floor ABCD = _____ m²

Calculate the length of EF, one of the horizontal edges of the block.

The length of EF = _____ m

This question carries only two marks, one for each part. In the first part, the pupil is required to read off the edge length of the 'floor' from the diagram, and to calculate the area. In the second, the edge length EF must be found. The real challenge is in identifying the right parts of the diagram from the description given, as the computation involved is simply halving.

6.1.2 TIMSS Mathematics Questions

The tests in the TIMSS surveys of 1999 and 2003 had a much greater proportion of multiple choice (over 50%), and a smaller proportion of constructed response questions than the PISA tests, but a similar range of question types was in evidence. For example, a question using a sequence of patterns with an identical structure to *Step Pattern* appeared in the 1999 cycle of TIMSS. This again requires pupils first to simply continue the number series.

TIMSS, MO22262, 1999 only

The figures show four sets consisting of circles.

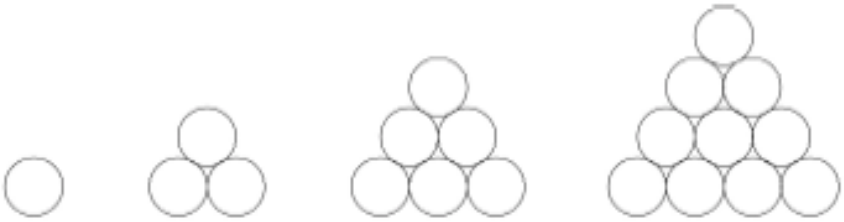


Figure 1 Figure 2 Figure 3 Figure 4

a) Complete the table below. First, fill in how many circles make up Figure 4. Then, find the number of circles that would be needed for the 5th figure if the sequence of figures is extended.

Figure	Number of circles
1	1
2	3
3	6
4	
5	

b) The sequence of figures is extended to the 7th figure. How many circles would be needed for Figure 7?

Answer: _____

In the last part of the question pupils are required to work with the relationship between the figure number and the number of circles.

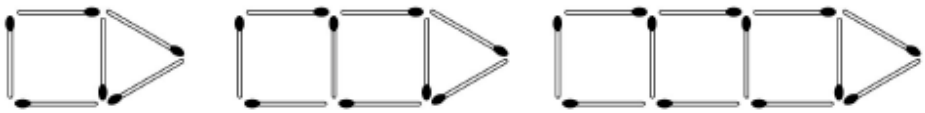
c) The 50th figure in the sequence contains 1275 circles. Determine the number of circles in the 51st figure. Without drawing the 51st figure, explain or show how you arrived at your answer.

Thus this four-mark question shows a significant similarity to the 2000 PISA question, *Apples*, although it focuses on the sequence of patterns found in the 2003 PISA question, *Step Pattern*. There is less reading than in the PISA item.

There is a much higher proportion of multiple choice questions in TIMSS than in PISA, including for example the question about a sequence of patterns of matchsticks shown below, which was used in both 1999 and 2003. Pupils are required to find the number of match sticks in the 10th member of the sequence, but they are not led through the initial stages of the reasoning (scaffolded) as they were in the question about the patterns of circles.

TIMSS, MO12017, 1999 and 2003

Matchsticks are arranged as shown in the figures.



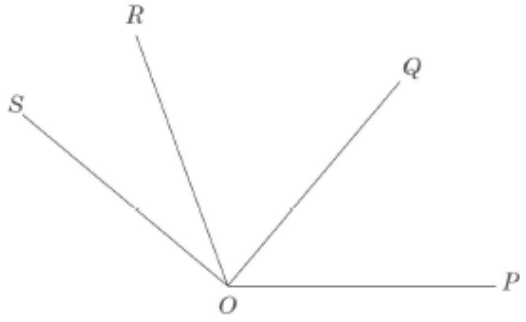
If the pattern is continued, how many matchsticks would be used to make Figure 10?

Ⓐ 30
Ⓑ 33
Ⓒ 36
Ⓓ 39
Ⓔ 42

Although many TIMSS one-mark questions are multiple choice, and there are fewer short response questions than in PISA, there are some examples of these. A 1999 question on angles, repeated in 2003, is of this type.

TIMSS, MO22202, ’99 & ‘03

In the figure, the measure of $\angle POR$ is 110°, the measure of $\angle QOS$ is 90°, and the measure of $\angle POS$ is 140°.



What is the measure of $\angle QOR$?

Answer: _____

As a short response question, this has a structure like that of the PISA question *Step Pattern*. However, in the topic assessed it is perhaps more similar to the PISA question *Triangles*.

Another type of short response question that is found in TIMSS has no parallels in PISA. This is a straightforward computation, without any context. For example, the 1999 cycle of TIMSS contained the following item on division.

Divide: $\frac{6}{55} \div \frac{3}{25} =$

Answer: _____

This question carries only one mark, with credit being awarded for the simplified response, $^{10}/_{11}$, or for any equivalent fraction.

We have noted that the 1999 cycle of TIMSS did include some multi-mark questions, such as the one about the sequence of patterns of circles. What is perhaps missing from TIMSS are questions like *Farm* in PISA, that carry just one or two marks but are presented with a great deal of extraneous detail. The 2003 cycle of TIMSS saw the introduction of a small number of more complex questions, in which a lot of data is presented from which the pupil must select what is needed to answer each part of the question. However, these questions carry several marks, and much more use is made, in the question as a whole, of the information given. One released example of this type of question is *Phone Plans*, shown below.

TIMSS, **Phone Plans** MO32762 to MO32764, '03

Betty, Frank, and Darlene have just moved to Zedland. They each need to get phone service. They received the following information from the telephone company about the two different phone plans it offers.

They must pay a set fee each month and there are different rates for each minute they talk. These rates depend on the time of the day or night they use the phone, and on which payment plan they choose. Both plans include time for which phone calls are free. Details of the two plans are shown in the table below.

Plan	Monthly Fee	Rate per minute		Free minutes per month
		Day (8 am – 6 pm)	Night (6 pm – 8 am)	
Plan A	20 zeds	3 zeds	1 zed	180
Plan B	15 zeds	2 zeds	2 zeds	120

Betty talks for less than 2 hours per month. Which plan would be less expensive for her?

Less expensive plan _____

Explain your answer in terms of both the monthly fee and free minutes.

Frank talks for 5 hours per month at the night rate. What would each plan cost him per month? Show your work.

Cost Per Month for Plan A: _____ zeds

Cost Per Month for Plan B: _____ zeds

Darlene signed up for the *Plan B*, and the cost of one month of service was 75 zeds. How many minutes did she talk that month? Show your work.

Minutes talked _____

This question has a longer introduction than most TIMSS items, which makes it superficially similar to some PISA questions. However, it has three parts, carrying two marks each, growing out of the context. It makes extensive use overall of the information given in the table at the beginning of the question, with much less redundancy than is found in PISA questions such as *Farm*. Even the longer TIMSS questions which were introduced in 2003 involve the student in taking in less information per mark than does a typical PISA question.

6.1.3 Key Stage 3 Mathematics Test Questions

At KS3 students take two written tests, one with a calculator and one without, and a mental mathematics test. The latter is unlike anything included at GCSE or in international surveys and is not discussed further here.

The KS3 tests contain a mixture of question lengths. There are a few short 1 mark questions, questions carrying several marks and taking up half a page, full page questions and some that are presented on a double page spread.

The reading demand of KS3 is low compared with PISA or TIMSS, and the complex sentence structures noted in the international surveys were not found. Unlike PISA or TIMSS, the number of marks that can be gained by each response is indicated to the student.

We have seen that, while the proportions vary, both PISA and TIMSS include some multiple choice, some short response, and some longer constructed response questions. However, PISA has a number of questions which include a quantity of extraneous data, while TIMSS questions generally make use of all of the data or information given in the question. Again, TIMSS has a few of out-of-context computations, which are not found in PISA. In both these respects, the questions in

the Key Stage 3 tests are more like those in TIMSS than in PISA. The tests again cover a range of types of question, but the range is closer to that found in TIMSS.

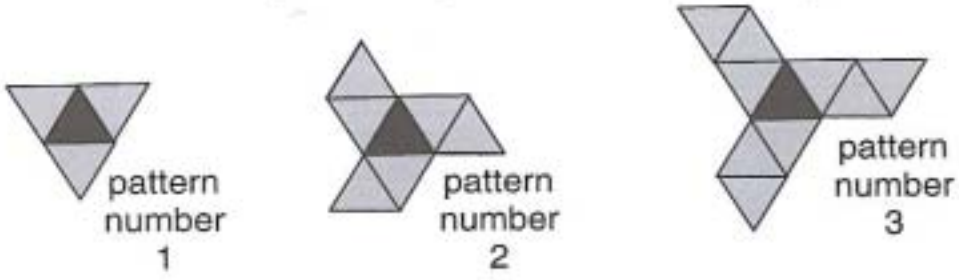
The 1999 Key Stage 3 question Arms, shown below, shows some similarities to the PISA and TIMSS questions on sequences of repeating patterns discussed above. However, in part (a) of this question, pupils are given the algebraic expression $1 + 3n$. They are required to relate the different parts of the expression to specific parts of the patterns. In this way, attention is drawn explicitly to the relationship between the patterns and the algebraic expression that represent them.

Parts b) and c) of the question are somewhat similar to the unscaffolded 'matchstick' question from TIMSS, but the pupils' attention has already been drawn to the structure of the patterns, rather than to the numbers per se.

Parts d) and e), however, are dissimilar to both the TIMSS and the PISA questions. A new pattern is introduced, and pupils are required to write their own algebraic expression, being guided clearly to work from the structure of the pattern in order to do so. Finally, in part e), the whole situation is reversed: pupils are given the expression, and are asked to develop a pattern to match it.

7. *Arms*

Jeff makes a sequence of patterns with black and grey triangular tiles.



The rule for finding the number of tiles in pattern number N in Jeff's sequence is:

number of tiles = **1 + 3N**

(a) The **1** in this rule represents the **black tile**.
What does the **3N** represent?

.....
1 mark

(b) Jeff makes **pattern number 12** in his sequence.
How many **black** tiles and how many **grey** tiles does he use?

.....
1 mark

..... black and ..... grey tiles

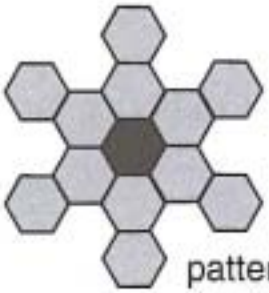
(c) Jeff uses **61 tiles** altogether to make a pattern in his sequence.
What is the number of the pattern he makes?

.....
1 mark

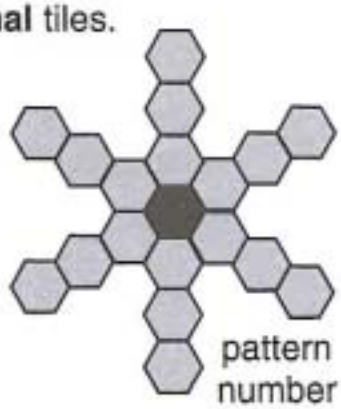
pattern number .....

(d) Barbara makes a sequence of patterns with **hexagonal** tiles.

pattern number 1



pattern number 2



pattern number 3

Each pattern in Barbara's sequence has **1 black** tile in the middle.

Each new pattern has **6 more grey** tiles than the pattern before.

Write the rule for finding the number of tiles in pattern number N in Barbara's sequence.

number of tiles = ... + ...

.... 1 mark

(e) Gwenno uses some tiles to make a **different** sequence of patterns.

The rule for finding the number of tiles in pattern number N in Gwenno's sequence is:

number of tiles = **1 + 4N**

Draw what you think the first 3 patterns in Gwenno's sequence could be.

This multi-mark question has much in common with the 2000 PISA and the 1999 TIMSS questions on the same topic, but the focus is on drawing the algebraic structure directly out of the sequence of patterns without the scaffolding provided by the tables of number patterns found in the PISA and TIMSS questions.

Whereas the great majority of TIMSS questions, and a number of PISA questions, carry only one mark, few of the KS3 mathematics test questions do so. Even where a computation is presented out of context, at least two marks are usually awarded. For example, the two-mark question below requires pupils to simplify their answer in order to obtain full credit. This is an example of a computation not placed in context. These are not uncommon in the KS3 tests, as are similar questions requiring algebraic manipulation.

KS3, 2004, 4-6 P1, q 19

Calculate $\frac{5}{6} \times \frac{3}{5}$

Show your working.

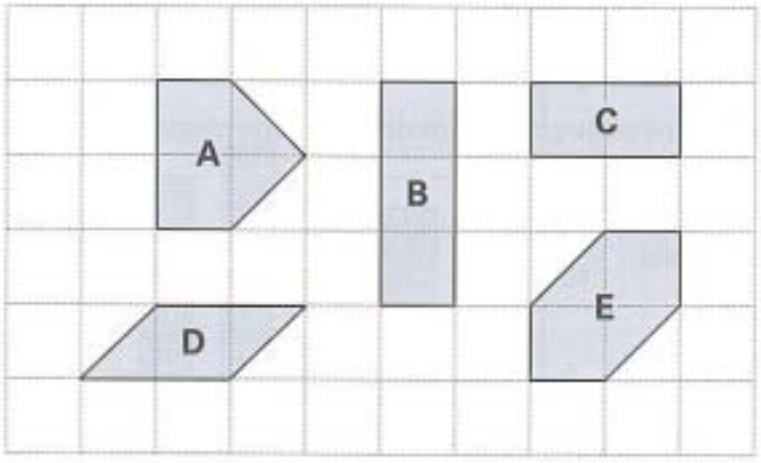
Write your answer as a fraction in its **simplest form**.

This question may be compared with the TIMSS computation with fractions shown above, which carried only one mark for which a simplified answer was not required.

Like PISA, the KS3 mathematics tests have a smaller proportion of multiple-choice type questions than TIMSS. Where they do occur, they usually form part of a longer question that exploits the same context. For example, parts (a) and (b) of the following question are a form of multiple choice, but part (c) is short-response.

KS3, 2004, 4-6 P1, q6

The diagram shows some shapes on a 10 by 6 square grid.



(a) Which **two** shapes have the **same area** as shape **A**?

.....

(b) Which **two** shapes have the **same perimeter** as shape **A**?

.....

(c) How many of shape **C** would you need to cover a 10 by 6 square grid?

.....

The Key Stage 3 mathematics tests do have a few single-mark, short response questions, such as the one shown below.

KS3, 2004, 4-6 p2, q 12

A box contains bags of crisps.

Each bag of crisps weighs **25 grams**.

Altogether, the bags of crisps inside the box weigh **1 kilogram**.

How many bags of crisps are inside the box?

.....

However, there are fewer of these than in PISA, and far fewer than in TIMSS. A higher proportion of the Key Stage 3 questions, having presented a context, then exploit that context for two or more marks.

6.1.4 GCSE Mathematics Questions

GCSE mathematics examinations involve the same range of question types as KS3 mathematics tests. A double page spread is the maximum noted, but many questions involve one page with a range of half page and shorter questions also used. Compared with KS3, many more of the one page questions comprise unrelated tasks that could be presented separately and in practice the GCSE papers can be seen as containing a high proportion of separate, single-mark items. An example is shown in the second example question below.

As for KS3, but unlike the international surveys, the number of marks that can be gained by a response is indicated to the student. GCSE questions responses carrying,

for example, five marks whereas two marks is now the usual maximum for KS3 and for TIMSS. Some questions are set in context, while others are not. Like KS3, the reading demand is low and complex sentence structures were not found.

The first question shown illustrates a one page GCSE question on number patterns. Part (a) requires the continuation of a simple pattern (this is the easiest tier of GCSE) but an explanation, such as adding 4, is also required. Part (b) requires the next pattern in the series to be considered and the use of a given rule. The second question, again one page, illustrates a more demanding task involving number patterns and an unrelated part (b) requiring the student to use trial and improvement.

The third question illustrates a geometry question requiring the use of Pythagoras' theorem in part (a), similar triangles in part (b) and tangent in part (c). (The values for some trigonometrical ratios are given because this is a non-calculator paper.)

Finally an example of a short response item is given. This carries 2 marks, with the possibility of one mark for a correct method but incorrect answer.

GCSE: SEG Foundation Tier 2002, Paper 5

5 (a) A sequence of numbers is shown.

8, 12, 16, 20, ...

(i) What is the next number in the sequence?

.....

.....

Answer ..... *(1 mark)*

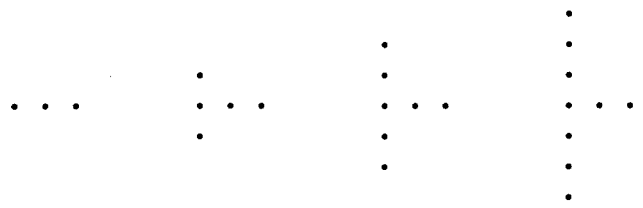
(ii) Explain how you found your answer to part (i).

.....

.....

(1 mark)

(b) Dots are used to make a sequence of patterns.



Pattern 1 Pattern 2 Pattern 3 Pattern 4

(i) How many dots will Pattern 5 have?

.....

.....

Answer ..... *(1 mark)*

(ii) The rule to work out the number of dots, d, for pattern number, p, is

$$d = 2p + 1.$$

Use the rule to work out the number of dots in Pattern 10.

.....

.....

Answer ..... *(2 marks)*

GCSE: SEG Intermediate Tier 2002, Paper 8

11 (a) (i) A sequence begins

$$-2, \quad -1, \quad 0, \quad 1, \quad 2, \quad \ldots\ldots$$

Write an expression in terms of n for the nth term of the sequence.

.....

.....

.....

Answer ..... *(2 marks)*

(ii) The nth term of a different sequence is $3n + 1$.

What is the difference between the 1st term and the 2nd term of the sequence?

.....

.....

.....

Answer ..... *(1 mark)*

(b) Use a trial and improvement method to find a solution to the equation

$$x^3 - 5x = 77.$$

Give your answer correct to one decimal place.
You **must** show all your working.

.....

.....

.....

.....

.....

.....

.....

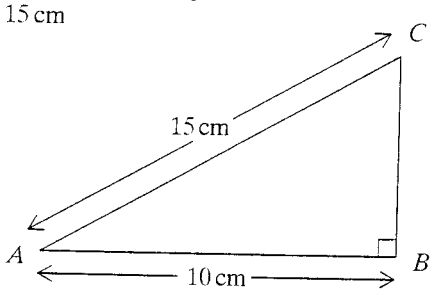
.....

.....

Answer $x =$ ..... *(4 marks)*

GCSE: AQA Higher Tier 2004, Paper 1

9 (a) The diagram shows a right-angled triangle ABC.
$AB = 10\,\text{cm}$ and $AC = 15\,\text{cm}$



Not drawn accurately

Calculate the length of BC.
Leave your answer as a square root.

.....

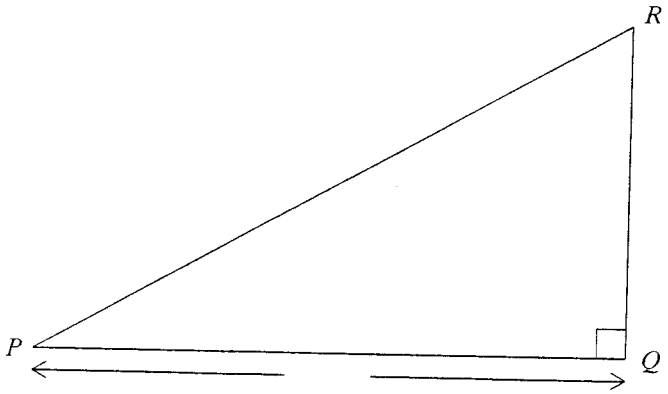
.....

.....

.....

Answer ..... cm *(3 marks)*

(b) Triangle PQR is similar to triangle ABC.
Angle CAB = angle RPQ.
$PQ = 25\,\text{cm}$



Not to scale

Work out the length of PR.

.....

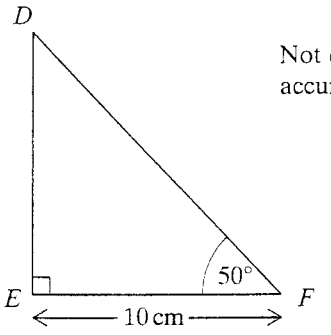
.....

.....

.....

Answer ..... cm *(3 marks)*

(c) The diagram shows a right-angled triangle DEF.
$EF = 10\,\text{cm}$
Angle $F = 50°$



Not drawn accurately

Angle	Sine	Cosine	Tangent
40°	0.643	0.766	0.839
50°	0.766	0.643	1.192

Use the table of data to work out the length of DE.

.....

.....

.....

.....

Answer ..... cm *(3 marks)*

GCSE: AQA Higher Tier 2004, Paper 1

6 Make p the subject of the formula $t = 5p + 40$

.....

.....

.....

Answer p = ... *(2 marks)*

6.1.5 Summary on Mathematics Questions

The comparisons between PISA, TIMSS, KS3 mathematics tests and GCSE mathematics can be summarized as follows.

Similarities and Differences between the Mathematics Assessments

	PISA	TIMSS	KS3	GCSE
In context	All questions	Some questions	Some questions	Some questions
Amount of Reading	Often Higher	Low	Low	Low
Complex sentence structures	Yes	Yes	No	No
2 page questions	Some	Some	Some	Some
Short questions	Some	Common	Some	Common
Multiple choice	Some	Over 50%	A few	A few

Both PISA and TIMSS show some differences from what students at KS3 and GCSE are familiar with. In PISA the placing of all questions in context is a significant difference from the other assessments, as is the amount of reading required. TIMSS and PISA both have complex sentence structures not found in KS3 or GCSE. There are variations in the proportions of multiple choice, short response and long question, but all the assessments essentially draw from the same pool of question types.

The familiarity rating exercise found TIMSS to be likely to be more familiar to students in England than PISA and the sort of comparisons made here do not contradict this. In mathematics the differences between the assessments are ones of degree. The overall picture is complex, and the effects of the differences on performance are not clear. The work carried out provides no evidence that the greater use of context in PISA matches what students in England are used to.

6.2 Question Styles in the Science Tests

As was noted previously all of the mathematics assessments involved a range of question styles and lengths and differences were often a matter of balance. In the science tests, however, PISA, KS3 and, to a lesser extent, GCSE have consistently adopted particular styles of question, while in TIMSS a variety of types has been used and this range of types increased in 2003.

In order to facilitate comparisons between the four assessments we have selected, where, possible, questions which address similar areas of science.

6.2.1 PISA Science Items

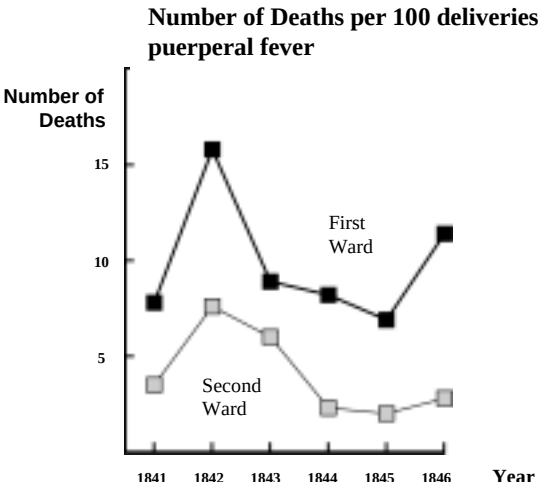
Only four science questions have been released from the PISA surveys, two from 2000 and two from 2003. However, these offer a valid sample of the PISA scientific literacy assessment items used in those surveys. Two of the questions, *Ozone* (2000) and *Daylight* (2003), were commented upon by the consultants, and these are discussed in that section of this report. The other two, *Semmelweis' Diary* (2000) and *Cloning* (2003) are given here.

All the PISA science questions are in many ways similar to the last type of PISA mathematics question discussed. A lot of information is presented, from which pupils need to select what is relevant while ignoring the rest. For example, *Semmelweis' Diary* is spread across three pages, and carries a total of six marks, with two constructed response parts carrying two marks each and two multiple choice parts.

PISA, SEMMELWEIS' DIARY TEXT 1

'*July 1846. Next week I will take up a position as "Herr Doktor" at the First Ward of the maternity clinic of the Vienna General Hospital. I was frightened when I heard about the percentage of patients who die in this clinic. This month not less than 36 of the 208 mothers died there, all from puerperal fever. Giving birth to a child is as dangerous as first-degree pneumonia.'*

These lines from the diary of Ignaz Semmelweis (1818-1865) illustrate the devastating effects of puerperal fever, a contagious disease that killed many women after childbirth. Semmelweis collected data about the number of deaths from puerperal fever in both the First and the Second Wards (see diagram).



Diagram

Physicians, among them Semmelweis, were completely in the dark about the cause of puerperal fever. Semmelweis' diary again:

'December 1846. Why do so many women die from this fever after giving birth without any problems? For centuries science has told us that it is an invisible epidemic that kills mothers. Causes may be changes in the air or some extraterrestrial influence or a movement of the earth itself, an earthquake.'

Nowadays not many people would consider extraterrestrial influence or an earthquake as possible causes of fever. We now know it has to do with hygienic conditions. But in the time Semmelweis lived, many people, even scientists, did! However, Semmelweis knew that it was unlikely that fever could be caused by extraterrestrial influence or an earthquake. He pointed at the data he collected (see diagram) and used this to try to persuade his colleagues.

Suppose you were Semmelweis. Give a reason (based on the data Semmelweis collected) why puerperal fever is unlikely to be caused by earthquakes.

.....

.....

.....

.....

PISA, SEMMELWEIS' DIARY TEXT 2

Part of the research in the hospital was dissection. The body of a deceased person was cut open to find a cause of death. Semmelweis recorded that the students working on the First ward usually took part in dissections on women who died the previous day, before they examined women who had just given birth. They did not pay much attention to cleaning themselves after the dissections. Some were even proud of the fact that you could tell by their smell that they had been working in the mortuary, as this showed how industrious they were!

One of Semmelweis' friends died after having cut himself during such a dissection. Dissection of his body showed he had the same symptoms as mothers who died from puerperal fever. This gave Semmelweis a new idea.

Semmelweis' new idea had to do with the high percentage of women dying in the maternity wards and the students' behaviour.

What was this idea?

A Having students clean themselves after dissections should lead to a decrease of puerperal fever.
B Students should not take part in dissections because they may cut themselves.
C Students smell because they do not clean themselves after a dissection.
D Students want to show that they are industrious, which makes them careless when they examine the women.

Semmelweis succeeded in his attempts to reduce the number of deaths due to puerperal fever. But puerperal fever even today remains a disease that is difficult to eliminate.

Fevers that are difficult to cure are still a problem in hospitals. Many routine measures serve to control this problem. Among those measures are washing sheets at high temperatures.

Explain why high temperature (while washing sheets) helps to reduce the risk that patients will contract a fever.

.....

.....

Many diseases may be cured by using antibiotics. However, the success of some antibiotics against puerperal fever has diminished in recent years.

What is the reason for this?

A Once produced, antibiotics gradually lose their activity.
B Bacteria become resistant to antibiotics.
C These antibiotics only help against puerperal fever, but not against other diseases.
D The need for these antibiotics has been reduced because public health conditions have improved considerably in recent years.

Semmelweis' Diary thus comprises a mixture of open response and multiple choice questions, offering to a total of 6 marks. *Cloning*, on the other hand, carries only 3 marks, one for each part, all of which are multiple choice.

PISA, CLONING

Read the newspaper article and answer the questions that follow.

A copying machine for living beings?

Without any doubt, if there had been elections for the animal of the year 1997, Dolly would have been the winner! Dolly is a Scottish sheep that you see in the photo. But Dolly is not just a simple sheep. She is a clone of another sheep. A clone means: a copy. Cloning means copying 'from a single master copy'. Scientists succeeded in creating a sheep (Dolly) that is identical to a sheep that functioned as a 'master copy'.

It was the Scottish scientist Ian Wilmut who designed the 'copying machine' for sheep. He took a very small piece from the udder of an adult sheep (sheep 1). From that small piece he removed the nucleus, then he transferred the nucleus into the egg-cell of another (female) sheep (sheep 2). But first he removed from that egg-cell all the material that would have determined sheep 2 characteristics in a lamb produced from that egg-cell. Ian Wilmut implanted the manipulated egg-cell of sheep 2 into yet another (female) sheep (sheep 3). Sheep 3 became pregnant and had a lamb: Dolly.

Some scientists think that within a few years it will be possible to clone people as well. But many governments have already decided to forbid cloning of people by law.

Question 1: CLONING

Which sheep is Dolly identical to?

A Sheep 1
B Sheep 2
C Sheep 3
D Dolly's father

Question 2: CLONING

In line 14 the part of the udder that was used is described as "a very small piece". From the article text you can work out what is meant by "a very small piece".

That "very small piece" is

A a cell.
B a gene.
C a cell nucleus.
D a chromosome.

Question 3: CLONING S128Q03

In the last sentence of the article it is stated that many governments have already decided to forbid cloning of people by law.

Two possible reasons for this decision are mentioned below.

Are these reasons scientific reasons?

Circle either "Yes" or "No" for each.

Reason:	Scientific?
Cloned people could be more sensitive to certain diseases than normal people.	Yes / No
People should not take over the role of a Creator.	Yes / No

Like all the science items in the PISA tests, both *Cloning* and *Semmelweis' Diary* require pupils to read a lot of material and then to extract what is needed to answer the questions. Unlike Reading literacy and Mathematics literacy, Science literacy has not yet formed a main focus of a PISA study, so there are fewer items, both released and unreleased, to compare. However, there is enough similarity between the items that are available to indicate that it is reasonable to speak of a 'PISA style' for the science questions, although it was not for the mathematics. This 'PISA style' is applicable to both the 2000 and the 2003 items, all of which require pupils to read through a lot of information, and then to make a selection to answer the questions.

6.2.2 TIMSS Science Items

While the PISA science items show a significantly narrower range of types of structure than the PISA mathematics items, in TIMSS the range is similar across the two subject areas. As for mathematics, around half of the TIMSS science questions are multiple choice, and carry only one mark. However, there are also a number of one-mark short response items, and a few extended questions that carry several marks all growing out of a given context. There is no clear equivalent to an uncontextualised mathematical computation in science, but the TIMSS science tests do include a number of items, such as the one shown below, that require knowledge of a simple scientific fact.

TIMSS, 2003, SO22206,

When chlorine gas reacts with sodium metal, what type of substance is formed?

Ⓐ A mixture
Ⓑ A compound
Ⓒ An element
Ⓓ An alloy
Ⓔ A solution

Two TIMSS items that assess ‘Reproduction and heredity’ are shown below.

TIMSS, 2003, SO12026

A son can inherit traits

Ⓐ only from his father
Ⓑ only from his mother
Ⓒ from both his father and his mother
Ⓓ from either his father or his mother, but not from both

TIMSS, 2003, SO22117

Why would male insects be treated to prevent sperm production?

Ⓐ To increase the number of female insects
Ⓑ To reduce the total population of insects
Ⓒ To produce new species of insects
Ⓓ To prevent insects from mating

All these questions have a simple multiple choice structure. Others require a short written response. This may be for one mark, as in the two questions below.

TIMSS, 2003, SO22154

Salil went to school with a cold. Several days later, half of his classmates also had colds. What is one likely reason some classmates had colds but others did not?

TIMSS, 2003, SO22160

What is the advantage of having two ears to hear with rather than one ear?

These questions indicate the range of one-mark items to be found in the TIMSS science tests. In addition to these, there are a number of longer questions carrying two or more marks.

TIMSS, 2003, SO32202

The diagram above shows a community consisting of mice, snakes and wheat plants.

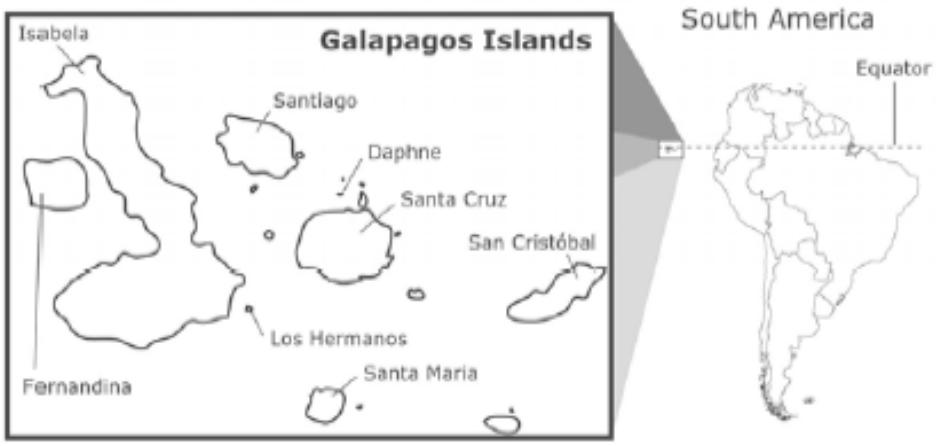
What would happen to this community if people killed the snakes?

Here, two marks are available for a response that mentions both that the number of mice would increase, and that the amount of wheat would decrease.

The TIMSS tests also included, from 2003, a small number of more extended science questions, somewhat similar to the mathematics item, *Phone Plans*. *Galapagos* is an example of this type of question. In this multi- part question, pupils have first to give an explanation for one mark.

TIMSS, 2003, SO32704

The diagram shows the Galapagos Islands, a group of volcanic islands in the Pacific Ocean about 1000 kilometers from South America. When these islands first formed they consisted of lava only. Eventually, living organisms (plants and animals) inhabited the islands. These organisms arrived long before people settled on the islands.



Which organisms that live on land most likely inhabited the Galapagos Islands first?

(Check one box.)

☐ Land plants

☐ Land animals

Explain your answer.

The second part of the question requires two reasoned responses, worth one mark each.

TIMSS, 2003, SO32705A

When settlers came to live on the Galapagos Islands, they brought with them a number of new animals such as cats and goats. Write down one effect the introduction of cats and goats could have on the animals and plants already living on the islands.

A. One effect of **cats**:

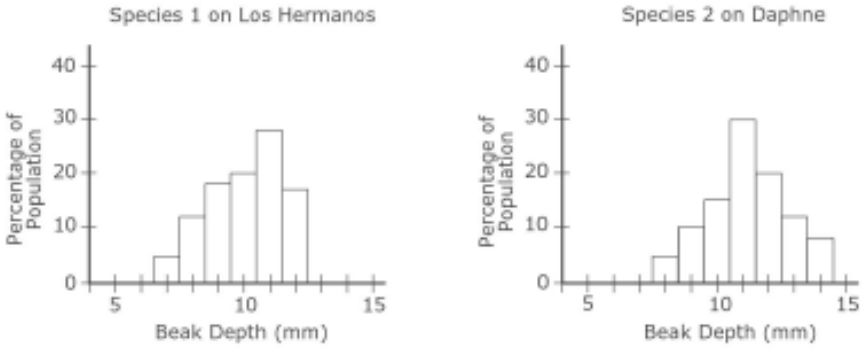
B. One effect of **goats**:

The third part of the question offers some graphical data, and requires pupils to interpret this to draw conclusions on two issues for one mark each.

TIMSS, 2003, SO32706

The Galapagos Islands contain a number of different species of finches (birds) that are thought to have developed from one species. Some species of finches eat certain types of seeds depending on their beak depth. The diagram below shows the head of one species of finch and its beak depth.

Some of the islands have only one species living on them, while other islands have more than one species. Species 1 lives on Los Hermanos Island. Species 2 lives on Daphne Island. The two graphs below show the percentage of the population with different beak depths for each of the two species.



A. How do the beak depths of Species 1 and Species 2 compare?

B. A wide variety of seeds exist on the islands, and both Species 1 and Species 2 eat seeds. Based on the beak depths of the two finch species, what would you conclude about the size of seeds that each species eats?

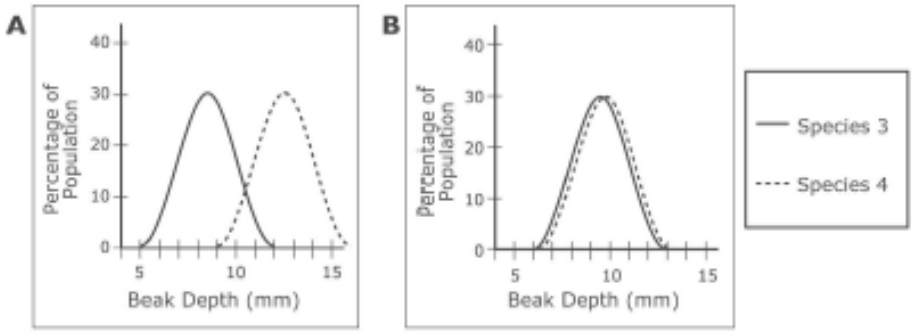
The final part of the question shows some related information in a different graphical format, and requires pupils to draw a conclusion relating to a different scenario, for one mark.

TIMSS, 2003, SO32707

Two other species (Species 3 and Species 4) live on Santa Maria Island, which also has a range of seed types.

Which of the following graphs shows a range of beak depths for Species 3 and Species 4 that would best insure the survival of both species on Santa Maria Island?

(Circle the letter by the correct graph.)



Explain why this range of beak depths would be best.

Thus this four part question carries six marks all together, one for each of six explanations or conclusions on different aspects of the situation described.

A question released in 1999 carried three marks, again offering one mark each for three different explanations or conclusions.

TIMSS 1999 SO22217A

A steel manufacturer uses a chemical process called 'galvanization' to protect the surface of steel beams that are used to construct high-rise buildings. It takes a crew of steel workers 8 hours to 'galvanize' a large batch of steel.

a) Why MUST the surface of the steel be protected?

b) A new 'galvanization' process is developed that shortens the procedure to 4 hours. Describe two consequences of the steel manufacturer switching to the shorter 'galvanization' process.

1.

2.

These two longer TIMSS questions show some similarities to the PISA style of science question, as they carry several marks and require pupils to draw conclusions, present hypotheses and construct quite complex explanations in their responses. However, even these do not require as much reading as the standard PISA science questions, and they constitute only a relatively small proportion of the set of TIMSS items as a whole.

6.2.3 The Key Stage 3 Science Tests

The format of questions in the KS3 science tests is relatively easy to characterise. Most are presented on two facing pages and contain a number of items addressing a common science theme or context. A few are presented on only one page, but the structure is similar. Questions tend to carry 5 or so marks. It should be noted that this sort of structure would have been familiar to students for some years, since the KS2 science tests are similar in structure but with rather more one page questions. Some examples are given below. Page layouts are quite spacious, and there is much less reading required than in PISA science items. The number of marks available for each response is always indicated. Because of the consistency of structure, only four examples are shown, two typical double page spreads and two sample single page questions.

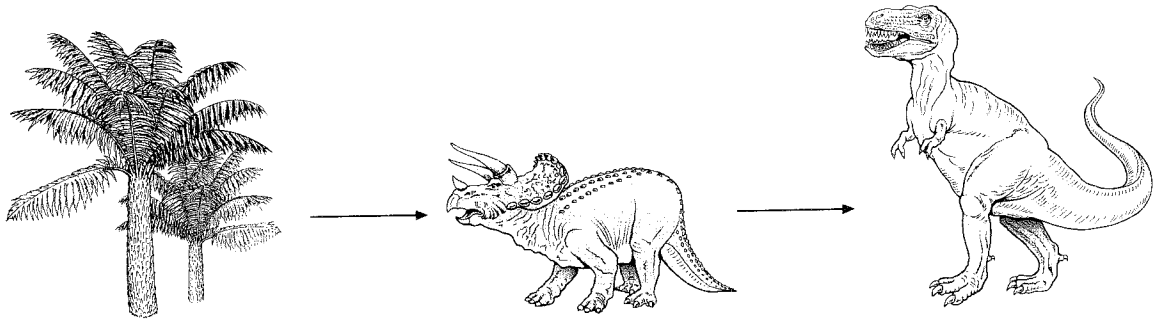
The first question concerns food chains, as did the TIMSS question on snakes, mice and wheat. The context here is the dinosaur era, and part (c) illustrates how multiple choice items are presented.

The second question, about scurvy, is, in some ways, similar to the PISA question *Semmelweis' Diary*. The amount of reading is unusually high for KS3, although it is still significantly lower than that found in PISA science questions. Such historically based contexts often require a larger number of words to set them up for the student, particularly when scientific investigations are involved.

The third question is an example of a single page presentation, and is another look at genetics and heredity. The context in this case is the extinct Quagga.

The fourth question, about acid rain, is set in a context whose social and political significance is comparable to that found in such PISA questions as *Cloning* and *Ozone*, discussed elsewhere. The scientific concept assessed in this question is the structure of a fair test, and in particular the control of variables. Questions that focus on this concept, and on the use of the terms *dependent* and *independent variable*, are common in national curriculum tests. There are questions with a similar focus, though without the terminology, in the unreleased parts of the international surveys.

9. The drawing shows a food chain including plants called tree ferns, and two dinosaurs.
They lived on Earth millions of years ago.



tree ferns Triceratops Tyrannosaurus rex

(a) The list below shows words which describe living things in a food chain.

herbivore **predator** **prey** **producer**

(i) Which word in the list above describes the tree fern?

1 mark

_____

(ii) From the list above, give **one** word that can describe Tyrannosaurus rex.

1 mark

_____

(iii) From the list above, give **one** word that can describe Triceratops.

1 mark

_____

(b) Some scientists think that a large rock from space hit the Earth about 65 million years ago.
A thick layer of dust stayed in the air for a long time and blocked out the sunlight.

This would cause a decrease in the number of tree ferns.
Give **one** way the decrease in tree ferns would affect the Triceratops.

1 mark

_____

_____

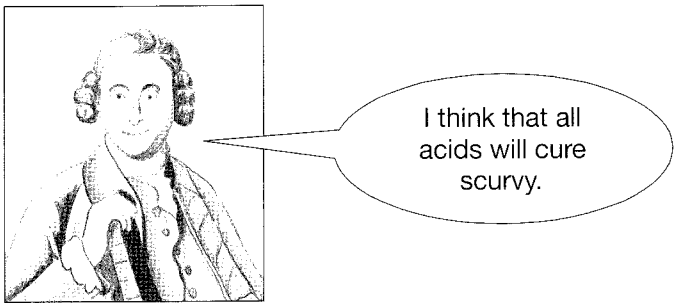
(c) Tyrannosaurus rex had thick scales covering its body.
Which group did it belong to?
Tick the correct box.

1 mark

amphibians	☐	fish	☐
reptiles	☐	mammals	☐

KS3, 2003, 3-6 P2, q 13

13. Sailors used to suffer from an illness called scurvy caused by a poor diet on long journeys.
James Lind was a doctor who tested treatments for scurvy. He predicted that **all acids cure scurvy**.



He gave 6 pairs of sailors with scurvy exactly the same meals but he also gave each pair a different addition to their diet.

pair of sailors	addition to their diet	effect after one week
1	some apple cider	beginning to recover
2	25 drops of very dilute sulphuric acid to gargle with*	still had scurvy
3	2 teaspoons of vinegar	still had scurvy
4	half a pint of sea water*	still had scurvy
5	2 oranges and 1 lemon	recovered
6	herbs and spices and acidified barley water	still had scurvy

(a) Does the evidence in the table support the prediction that all acids cure scurvy?
Tick the correct box.

☐ yes ☐ no

Use the table to explain your answer.

1 mark

_____

_____

***DANGER! DO NOT TRY THIS.**

(b) (i) Give the **one** factor James Lind **changed** in this experiment.
(This is called the independent variable.)

1 mark

(ii) Give the factor James Lind **examined** in this experiment.
(This is called the dependent variable.)

1 mark

(c) James Lind's evidence suggested that oranges and lemons cured scurvy.

At a later time, other scientists did the following:

- They separated citric acid from the fruit.
- They predicted that citric acid would cure scurvy.
- They tested their prediction by giving pure citric acid as an addition to the diet of sailors with scurvy.
- They found it did **not** cure scurvy.

The scientists had to make a different prediction.

Suggest a new prediction about a cure for scurvy that is consistent with the evidence collected.

1 mark

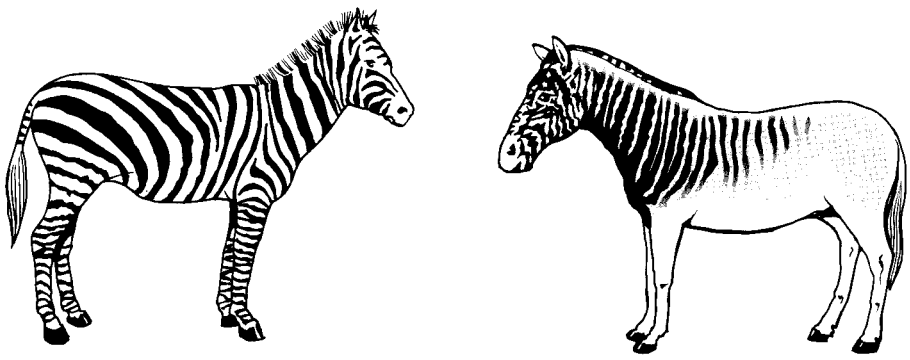
(d) Explain why it is necessary to investigate the effects of changes in diet over a period of more than one week.

1 mark

maximum 5 marks

KS3, 2003, 5-7 P1, q 10

10. The quagga is an extinct animal that lived in Africa. Quaggas belonged to the same group as zebras.
The drawings below show a zebra and a quagga.



zebra

quagga

(a) Zebras and quaggas used to breed with each other. The offspring contained a combination of both zebra and quagga genes (genetic information).
How were zebra **and** quagga genes passed on from the parents to their offspring?

1 mark

(b) These days there are some zebras that still show some quagga features.
Scientists are using zebras to try to produce quaggas by selective breeding.
Describe the steps in this selective breeding process.

3 marks

maximum 4 marks

KS3, 2003, 5-7 P2, q 12

12. A headline from a newspaper is shown below.

Some countries claim that acid rain caused by power stations in Britain damages their forests.
Others argue that coal-burning power stations produce cheap electricity and that plants can stand some level of acid rain.

Imagine you are planning a laboratory investigation of the claim:

'plants can stand some level of acid rain'.

Assume you have access to whatever laboratory equipment you need, including:

- seeds
- acid
- seed trays
- soil

Plan a laboratory investigation to test the claim that **'plants can stand some level of acid rain'**.

(a) Name a factor you would need to vary in your investigation.
(This is the independent variable.)

1 mark

(b) (i) What factor would you examine to see the effect?
(This is the dependent variable.)

1 mark

(ii) How could you measure this dependent variable?

1 mark

(c) Suggest **one** factor you would control to ensure that your investigation is fair.

1 mark

maximum 4 marks

6.2.4 GCSE Science Examinations

GCSE Science Tests

As for the KS3 tests, the format of questions in the GCSE papers we have examined is very consistent. Some are presented on two facing pages and contain a number of items addressing a common science theme. There are rather more one page questions than at KS3, but again the structure is similar to that for the two page questions. Unlike the KS3 tests, there are also short questions, say half a page or so. While many questions, like KS3, carry around five marks there are a number with more than this. A difference from KS3 is that a there are fewer items carrying only 1 mark. Again the number of marks available for each response is clearly indicated. Some examples are shown below. The tier from which the question is taken is indicated, and, again, the range used, from double page to short item is illustrated.

The first question, a single page one, takes cloning as its subject, as in the PISA item, *Cloning*. The second question, also one page, looks at an environmental issue, the possible dangers arising from the over-use of antibiotics.

The third question, on a double page spread, presents a large quantity of data on the effect of light on plant growth, and asks for patterns to be found and described. The student is required to use the information in order to construct a good, reasoned response. However, although a lot of data is provided, little of it is actually redundant: the challenge is to sort through and organise the information effectively, rather than to reject much of it as irrelevant.

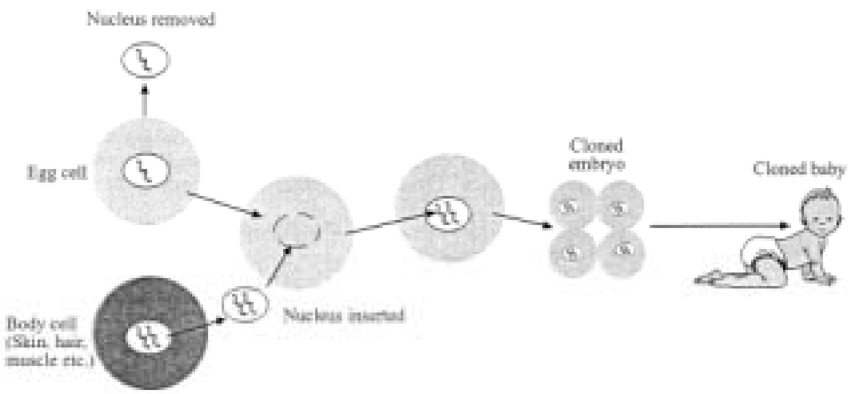
The last question again presents data, but asks for a graph to be completed. It then goes on to ask for interpretations of the situation and is completed by a matching item.

GCSE: **AQA Higher Tier 2003,**

Double science award (modular), Paper 1, Question 2

LEAVE MARGIN BLANK

2 It is now possible to clone humans. The diagram shows one way in which this can be done.



(a) What type of reproduction is this?

....

(1 mark)

(b) Will the baby have the characteristics of the egg cell or the body cell?

....

Explain the reason for your answer.

....

....

....

....

(2 marks)

(c) The procedure in the diagram could be used to produce several cloned embryos.

Suggest how this might be done.

....

....

(1 mark)

4

Turn over ▶

G/H134859/S03/3466/H1

GCSE: AQA Higher Tier 2003,

Double science award (modular), Paper 1, Question 3

3 Read the passage about antibiotics.

People do not always agree about the use of antibiotics in food production.

If we put low doses of antibiotics in feed for animals such as cattle and sheep, it helps to produce high-quality, low-cost food. Antibiotics help to keep animals disease-free. They also help animals to grow. Animals get fatter quicker because they do not waste energy trying to overcome illness.

The use of antibiotics in livestock feed means that there is a higher risk of antibiotic-resistant bacteria developing. The rapid reproduction of bacteria means there is always a chance that a population of bacteria will develop which is antibiotic-resistant. These could be dangerous to human health.

(a) *To gain full marks for this question you should write your ideas in good English. Put them into a sensible order and use the correct scientific words.*

Explain how a population of antibiotic-resistant bacteria might develop from non-resistant bacteria.

.....

.....

.....

.....

.....

.....

(3 marks)

(b) Do you think that farmers should be allowed to put low doses of antibiotics in animal feed?

Explain the reasons for your answer.

.....

.....

.....

.....

(2 marks)

GCSE: AQA Higher Tier 2003,

Double science award (modular), Paper 1, Question 20

20 Low light intensity is one factor that limits the yield of a crop.

In Britain, many tomato growers use artificial lights to increase the yield of tomato crops.

The table shows the amount of natural daylight and artificial lamplight received by a tomato crop grown in a greenhouse.

Month	Natural daylight received by tomato plant		Artificial lamplight given to tomato plant		Total light energy received by plant per day in J/cm^2	Percentage increase in growth resulting from artificial light
	Day length in hours	Light energy received by plant per day in J/cm^2	Hours of light given per day	Light energy received by plant per day in J/cm^2		
January	8.1	239	18	492	731	206
February	9.9	492	18	492	984	100
March	11.9	848	12	328	1176	39
April	13.9	1401	2	55	1456	4
May	15.5	1786	0	0	1786	0
June	16.6	1960	0	0	1960	0
July	16.2	1849	0	0	1849	0
August	14.7	1561	0	0	1561	0
September	12.8	1064	2	55	1119	5
October	10.6	614	11	301	915	49
November	8.8	288	18	492	780	171
December	7.6	183	18	492	675	269

(a) Describe the pattern for the amount of light energy received from natural daylight by a tomato plant during the year.

.....

.....

.....

.....

.....

.....

(3 marks)

(b) A tomato plant needs 600 *J* of light energy per cm^2 each day to grow and produce tomatoes.

Use this information and data from the table to suggest an explanation for the pattern of the artificial light given to the tomato plants.

.....

.....

.....

.....

.....

.....

(2 marks)

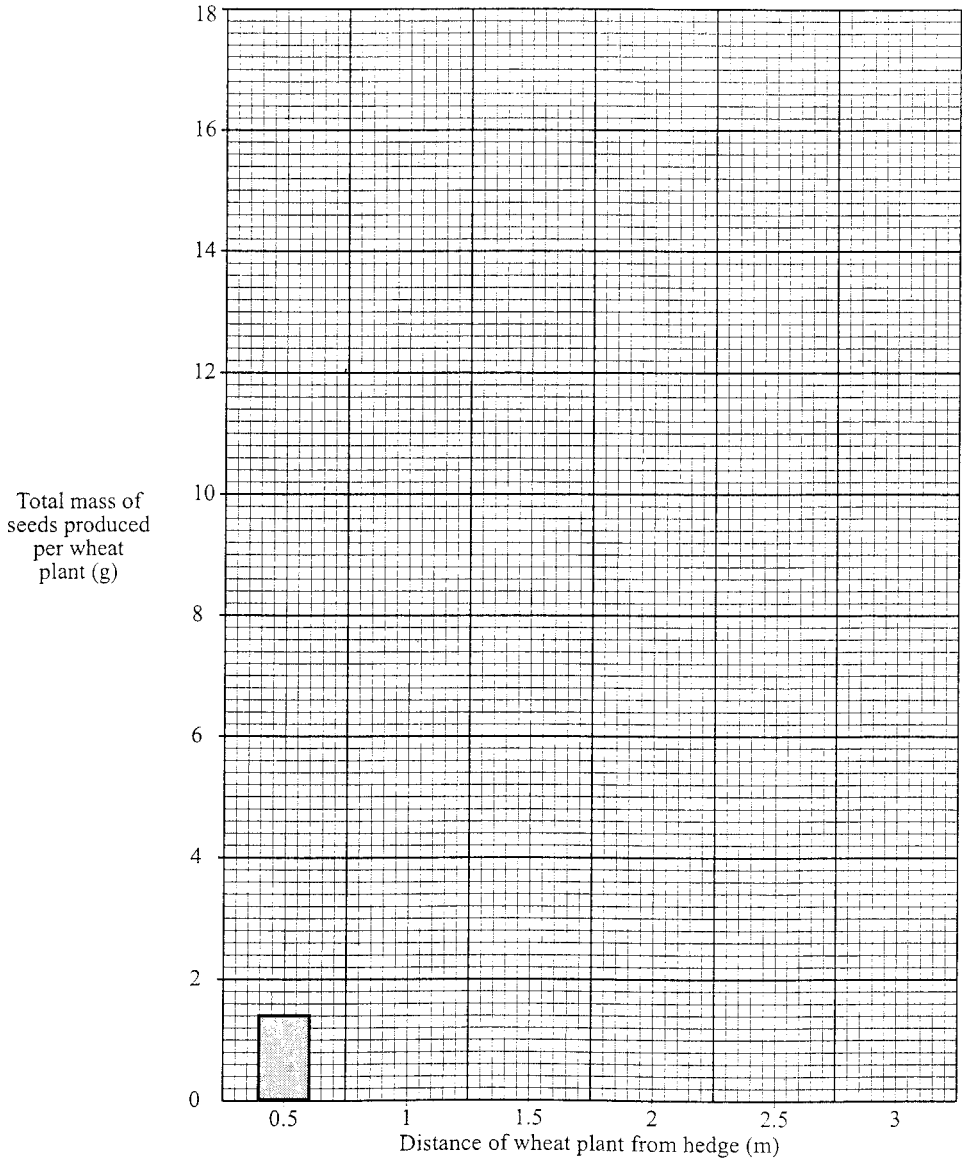
GCSE: AQA Foundation Tier 2001,

Double science award (co-ordinated), Paper 1, Question 10

10 A wheat field is surrounded by hedges. The effect of the hedge on the growth of wheat was investigated.
The total mass of seeds produced by wheat plants at different distances away from the hedge was measured.
The results are shown in the table.

Distance of wheat plant from hedge (m)	0.5	1.0	1.5	2.0	2.5	3.0
Total mass of seeds produced per wheat plant (g)	1.4	3.2	7.8	16.0	16.2	16.2

(a) Draw a bar chart to show the results of the investigation. The first bar has been drawn for you.



(2 marks)

(b) (i) Describe the effect of the hedge on seed production by the wheat plants.

.....

.....

.....

.....

(2 marks)

(ii) Suggest **one** reason why the hedge affects seed production by the wheat plants.

.....

.....

(1 mark)

(c) The farmer decides to cut down the hedges so that he can grow more wheat.
Soon he notices that fewer barn owls are living in the area.
Barn owls do **not** nest in hedges and do **not** eat the plants in the hedge.
Suggest **one** reason why there are fewer barn owls.

.....

.....

(1 mark)

(d) **List A** gives the names of some substances that may affect the environment.
List B gives information about these substances.

List A	**List B**
sewage	added to crops to promote growth
fertiliser	used for killing insects that eat crops
pesticide	lowers the oxygen content of rivers
herbicide	dissolves in rain water to produce acid rain
sulphur dioxide	sprayed on crops to kill weeds

Draw a straight line from each substance in **List A** to the information about it in **List B**.
One has been done for you.

(4 marks)

6.2.5 How PISA and TIMSS Science Items are likely to Come Across to English Students

As we have shown, the science items that English students will have seen and attempted in national tests and examinations are consistent in style. The path from KS2 to KS3 to GCSE is composed mainly of questions presented on one or two complete pages. In either case a scientific theme or context sustains a number of marks. Students taking PISA will also have seen shorter science questions in preparation for GCSE. The students taking TIMSS will have experienced shorter questions in a range of subjects, but the extent to which these will have included science will vary from school to school. (Commercially produced science tests are rare for this age range.)

A KS3 student meeting a TIMSS test will find comparatively few two page questions of the type with which they are familiar from KS3, but they may encounter some that spread over four pages or more. The short TIMSS items are likely to be relatively unfamiliar inasmuch as they address science, but their format is unlikely to be really novel, having been encountered in other subject contexts. This includes multiple choice items, much more common in TIMSS than in KS3 or GCSE but present in both in a variety of formats, although not the precise one used in TIMSS.

PISA science questions from both 2000 and 2003 are likely to be unfamiliar to English students because of the amount of reading required and their rather different structure. This difference arises mainly from PISA's focus on scientific literacy. The emphasis on real-world contexts almost inevitably requires a fair amount of text both to set up the context and to give other information. However, as has been shown, similarly contextualised questions can be found in both KS3 and GCSE but with significantly less reading required.

Both TIMSS and PISA are likely to present science to English students in ways with which they are unfamiliar, or relatively so. The likely effects on performance are more difficult to judge, particularly in TIMSS where a variety of formats are used. In PISA, the willingness of students to read the required amount of text, and their ability to do so, are likely to be the crucial factors.

The rating exercise found no overall differences between PISA and TIMSS items in likely appropriateness or familiarity to English students. Looking at the wider structure of the papers and questions has identified differences, particularly with regard to the amount of reading and the need to deal with extraneous information that are characteristic of PISA, but no clear conclusions may be drawn.

6.3 PISA Reading Literacy, KS3 English and GCSE Compared

Since there is no international study to compare PISA reading literacy with, this section examines the relationship between three different contexts for the assessment of pupils' reading literacy in Years 9-11: PISA, KS3 tests and GCSE. The following questions represent the main focus for this investigation:

- How far does PISA reading reflect what we call school subject English?
- On what basis can we say how familiar and accessible is PISA reading?
- How far are pupils in England called upon to offer an opinion in tests or exams?

The nature of both reading literacy and English differs from both mathematics and science, and the general assessment focuses in the three assessments are compared, and the text types used are analysed.

The following sections are based on a scrutiny of PISA reading items but direct reference is made in this report only to those items already released into the public domain. Approaches to the assessment of reading at KS3 and GCSE are described and a range of representative KS3 and GCSE questions is given. These sections incorporate comparative commentaries on similarities and differences between PISA and KS3/GCSE questions. In the final section, direct answers are offered to the key questions listed above and to others which emerge during the course of the investigation.

6.3.1 PISA Reading Literacy

The OECD 2002 overview states that the purpose of PISA is to measure how well 15 year-olds are able to "meet the challenges of today's knowledge societies" by focusing on their ability to use their knowledge and skills to meet real-life challenges, rather than ... the extent to which they have mastered a specific school curriculum" (p9). The concept of real-life is reinforced on page 11 in a reference PISA's "more explicit focus on the knowledge, understanding and skills required for effective functioning in everyday life."

OECD argues that reading literacy processes move "beyond the ability to retrieve surface meaning from text" to "an understanding and appreciation of the writer's craft, and an ability to reason about text" and the application of "an understanding of text structure, genre, and register".

Although the references to "everyday life" do not have exact counterparts in the programmes of study for KS3 and KS4 or the GCSE specifications, the higher level reading processes referred to in the previous paragraph do have clear echoes in the KS3 reading assessment focuses and the GCSE reading assessment objectives which are outlined later. The PISA focus emphasises:

> "…the ability to stand apart from arguments, evidence or text, to reflect on these, and to evaluate and criticise claims made [and] skills [that] go beyond analysis, problem solving and communication to evaluation and critical reflection."

These too are readily compatible with many of the terms used to identify the range and hierarchy of reading skills which are taught and learnt at KS3 and KS4, and further consolidated and extended at AS/A2 level in English specifications.

Reading literacy is defined for PISA purposes as:

> *"The capacity to understand, use and reflect on written texts, in order to achieve one's goals, to develop one's knowledge and potential and to participate in society."*

This definition is further developed to identify PISA literacy processes:

> "Reading literacy:
>
> Performing different kinds of reading tasks such as forming a broad general understanding, retrieving specific information, developing an interpretation or reflecting on the content or form of the text."

The PISA literacy definition for the concept of knowledge and understanding provides more detailed exemplification:

> "Reading literacy
>
> *Reading different kinds of text: continuous text classified by type (e.g., description, narration, exposition, argumentation and instruction) and non-continuous texts classified by structure (e.g., forms, calls and advertisement, charts and graphs or tables)."*

PISA test items are classified according to the following criteria:

- Task
- Text
- Question and response formats.

Tasks are designed to test pupils' abilities in three aspects of reading performance:

- Interpreting texts
- Retrieving information
- Reflection and evaluation.

Texts used in PISA tests are categorised as either **non-continuous** or **continuous**. In practice, this means that any unit mainly consisting of data in the form of graphs, bar charts, tables, tree diagrams, flow charts, etc is classed as non-continuous and the various texts containing sustained prose are classed as continuous, even though a wide range of subjects, genres and styles is represented. Of the eight examples of continuous prose in the released items, six are non-literary and two are literary texts:

Unit 2: Flu	information sheet

Unit 3: Graffiti	two letters from an internet site
Unit 6: Police	magazine article
Unit 7: Runners	article
Unit 8: The Gift	short story
Unit 9: Amanda and the Duchess	(i) play script (extract) (ii) glossary (definitions of theatrical occupations)
Unit 10: Personnel	written announcement
Unit 11: New rules	newspaper editorial

The three released non-continuous items are:

Unit 1: Lake Chad	(i) graph (ii) bar chart
Unit 4: Labor	tree diagram
Unit 5: PLAN International	table

Question and response formats are of three types**:**

- Multiple choice response
- Short answer
- Extended response.

The first of these formats is used to a limited extent and in modified forms in KS3 English tests and the second and third are familiar at both KS3 and GCSE.

A Processes or reading tasks

1. forming a broad general understanding (20%)
2. retrieving information (20%)
3. developing an interpretation (30%)
4. reflecting on content (15%)
5. reflecting on form (15%)

These PISA processes and concepts are also relevant to assessment approaches at KS3 and GCSE, although some differences of nomenclature occur in the different contexts.

B Context or text format

1. **Continuous** – in standard 'prose' form

a descriptive: what? questions (13%)

b narrative: when?/in what sequence? questions (13%)

c expository: how? questions (22%)

d argumentative/persuasive: why? questions (13%)

e injunctive: directions (5%)

These PISA text type categories are also compatible with text types specified for the reading curriculum at KS3 and GCSE levels. However, there is one notable omission from the PISA list: poetic or verse forms, which feature heavily at KS3 (through the compulsory Shakespeare study) and also in GCSE specifications.

2. **Non-continuous** or **situation**

a charts/graphs (11%)

b tables (11%)

c diagrams (3%)

d maps (3%)

e forms (3%)

f advertisements (2%)

This final category (f) is perhaps problematic: the only PISA example which appears to fall within this category is a secure unit consisting of two texts, one being continuous free verse, yet set out with a formalised line structure, and the second an uncaptioned photograph. Taking forms (a) to (e), however, a significant mismatch between PISA expectations and those underlying the KS3/KS4 English curriculum becomes apparent. Many English teachers would argue that categories (a) to (e) are more appropriate to study in other curriculum areas, and the particular conventions of the presentation of numerical information in tabular or graphical form is not essentially part of the English reading curriculum. That is not to say that such skills should not be taught, or are unimportant, but it could be argued that these forms are more properly to be considered as the province of mathematics or geography.

The most likely context for encountering these forms within English would be in the context of media texts, but here the level of complexity and detail would generally fall far short of what is presented in the PISA non-continuous texts.

C Context or situation

1. **Personal (private):** carried out to satisfy an individual's own interests both practical and intellectual (28%)

2. **Educational:** reading to learn, normally involved with acquiring information as part of a larger learning task (28%)

3. **Occupational:** closely associated with the accomplishment of some immediate task (16%)

4. **Public:** carried out to participate in the activities of the larger society (28%)

These four categories do not fit comfortably with the English reading curriculum, governed as it is by the NC requirements to cover a broad literary range at KS3 and GCSE levels. At KS3 and KS4 the reading requirements are summarised as:

- ***Understanding texts***
- ***Reading for meaning***
- ***Understanding the author's craft***
- ***English literary heritage***
- ***Texts from different cultures and traditions***
- ***Printed and ICT-based information texts***
- ***Media and moving image texts.***

In England, GCSE specifications are required to assess a range of reading to include prose, poetry and drama. Within this, the range must contain:

- a play by Shakespeare
- work from the English literary heritage by at least one major writer with a well established critical reputation
- texts from different cultures and traditions
- non-fiction texts
- media texts.

None of the four PISA contexts appears to incorporate literary texts as an essential genre, although in practice some literary texts do appear in the released and secure items. Overall, however, the PISA reading diet seems to be biased in favour of transactional, non-literary texts of a kind addressing a predominantly adult readership. *The Gift* is a literary text of a kind likely to form part of the reading range for GCSE but it is presented in the PISA booklet in a very small point size and with long lines. Many students would have performed better with this text if it had been more sympathetically and appropriately presented.

Many of the PISA items are sufficiently similar to those encountered in KS3 English tests and GCSE examinations for students to recognise them as reasonably familiar. However, significant numbers of PISA stimulus texts are of a kind not deployed in English tests or GCSE examinations. These texts fall into the following main categories:

- those in which the subject matter is unfamiliar because it is outside the experience of most UK students within the target age range (e.g. Unit 11: Technology creates the need for new rules)

- those in which information is presented in tabular or graphical forms (e.g. Unit 1: Lake Chad; Unit 4: Labour; and Unit 5: PLAN International)

- those in which conceptual difficulties occur because secure comprehension is dependent upon prior knowledge (e.g. Unit 6: Police; Unit 11: New rules)

- those exhibiting complexities of syntax and/or vocabulary, or other linguistic difficulties (e.g. [some secure units])

- continuous texts in which line lengths exceed those encountered in book or magazine formats, thereby affecting readability (e.g. Unit 8: The Gift)

- those in which insufficient preliminary information is given to help readers to contextualise the reading matter (e.g. [some secure units])

[Note that complete PISA Reading Literacy questions are not included in this section of the report, for reasons of space. However, all released Reading Literacy items may be downloaded from: http://www.pisa.oecd.org/dataoecd/44/62/33692744.pdf]

All of the PISA texts are encountered by students as unseen, as is also the case with KS3 English tests. In the latter case, however, all questions are based on three substantial passages that are linked thematically. These three texts represent different genres and so provide a range of different reading experiences and contexts for question-setting. In 2003 the theme was *In Search of Treasure* and the first extract was taken from a non-fiction book about treasure-hunting, the second was an extract from Stevenson's *Treasure Island* and the third an extract from Howard Carter's diary in which he described his first sight of the treasure inside Tutankhamun's tomb. An environmental theme *Save It* was used in 2004; the first text was newspaper article about the work of bin men, the second was an extract about the effects of tourism from a text book and the third consisted of pages from the *Eden Project* website.

In all cases, the KS3 reading materials are of an appropriate level of difficulty and are presented attractively with full use of colour where appropriate. Font type and point size have also evidently been selected in order to enhance the accessibility of the texts. This is, to a lesser extent, also true of the GCSE reading texts but in general the presentation of the PISA items is less successful. The difficulties inherent in encountering unfamiliar text types without having been taught, for example, the presentational conventions of international or domestic passenger timetables, is bound to affect comprehension levels on such items. On the whole, these difficulties are avoided at KS3 and GCSE levels by the use of accessible text types for unseen examinations. At KS3 Shakespeare texts are studied and for GCSE a wider range of reading matter is specified. Literary texts are studied as part of the course but media and non-fiction texts tend to be encountered as unseen.

Reading comprehension is not solely related to the linguistic demands made by the text on readers. Texts dealing with familiar subject matter are perceived as easier to comprehend even though they may be as linguistically demanding as texts dealing with more obscure subject matter. One of the PISA texts most accessible to the target age range, Unit 7: Runners, contains challenging syntax and vocabulary but respondents achieved an 82% correct response rate to the four questions set on this text. The UK correct response rate was 83.25%. This PISA text and the associated questions make an interesting comparison with an Edexcel GCSE Foundation Tier

paper used in June 2003. The stimulus text was a *Daily Mirror* report on the launch of Nike Shox trainers, and the question was:

> Read the article about Nike Shox trainers.
>
> How does the article try to interest Daily Mirror readers in this subject?
>
> You should write about:
>
> - the content of the report
> - the choice of language
> - the headlines and captions
> - the use of photographs
> - any other features of layout and design.

This question provides an angle for the reader to consider (***How** does the article **try to interest** …readers*) but also uses bullet points to help students to focus and structure their written responses. Candidates are essentially required to select from a wide range of possible content points and to offer an individual interpretation of the text, according to the focus of the question. The mark scheme sets out broad criteria that distinguish between levels of evaluative response rather than a list of essential points that must be made or evidence that must be cited. This style of question typically elicits a continuous prose response of 2-3 A4 pages and contrasts with the PISA style of a larger number of more specifically focused questions:

Q45 What does the author intend to show in this text?

A That the quality of many sports shoes has greatly improved.

B That it is best not to play football if you are under 12 years of age.

C That young people are suffering more and more injuries due to their poor physical condition.

D That it is very important for young sports players to wear good sports shoes.

Q46 According to the article, why should sports shoes not be too rigid?

Q47 One part of the article says, “A good sports shoe should meet four criteria.” What are these criteria?

Q48 Look at this sentence from near the end of the article. It is presented here in two parts:

> “To avoid minor but painful conditions such as blisters or even splits or athlete’s foot (fungal infections),…”
>
> “…the shoe must allow evaporation of perspiration and must prevent outside dampness from getting in.”

What is the relationship between the first and second parts of the sentence?

> The second part
>
> A contradicts the second part.
>
> B repeats the first part
>
> C illustrates the problem described in the first part
>
> D gives the solution to the problem described in the first part.

The first of the PISA questions requires pupils to arrive at an interpretative overview of the whole text by considering authorial intention but the second and third questions target the retrieval of information. The final question targets reflection and evaluation by requiring pupils to consider the semantic and structural relationship between two adjoining clauses.

In addition to the difficulties inherent in some of the text types deployed in the PISA items, some of the questions are framed in unfamiliar ways unlikely to be encountered in English tests or GCSE examinations.

For example, the wording of some PISA questions may be unnecessarily obscure if interpolated content interrupts the completion of the main clause, e.g.:

> List two examples from the editorial that illustrates (sic) how modern technology, such as that used for implanting frozen embryos creates the need for new rules.

Some PISA questions are longer than is generally the case with KS3 and GCSE. The third question on *The Gift* is:

Here are some of the early references to the panther in the story.

> "the cry awoke her, a sound so anguished…" (line 24)
>
> "The answer was a repeated cry, but less shrill, tired sounding…"
>
> (line 33)
>
> "She had…heard their cries, like suffering in the distance."
>
> (line 39-40)

Considering what happens in the rest of the story, why do you think the writer chooses to introduce the panther with these descriptions?

There is also a further difficulty in this question - the potential for misunderstanding that comes from the possible implication that it is "the panther" in the first line of the question could be confused with the references to "her" and "She" in the quotations, which actually refer to the woman living in the house.

The 2002 overview of PISA places heavy emphasis on the utilitarian and transactional applications of reading literacy. There is a firm emphasis on such issues as "skills that are relevant to adult life" and "causes and consequences of observed skill shortages". Three of the four reading contexts specified also have a transactional emphasis:

- Reading for public use… (… carried out to participate in the activities of the wider society)
- Reading for work… (… to equip students for the world of work)
- Reading for education… (… normally involved with acquiring information as part of a larger learning task).

The remaining context, Reading for private use (personal), includes:

> "personal letters, fiction, biography and informational texts read as part of leisure or recreational activities".

This incorporates the categories of literary text specified for the KS3/KS4 reading curriculum but also includes various types of non-fiction texts. The released and secure PISA literary texts form a much smaller proportion of the total reading

curriculum than is the case at KS3/KS4. Furthermore, the literary texts which are featured are not, on the whole, as accessible as the unseen texts used in KS3 tests and GCSE examinations.

6.3.2 The KS3 English Tests

The KS2 and KS3 tests, as well as the optional Progress Tests intended for pupils between the ages of 9 and 14, are designed to test specific components of pupils' reading performance, and also to provide diagnostic evidence on individual pupils' performances and on the relative strengths and weaknesses of whole classes or sub-groups within classes or cohorts. In this, they have something in common with the PISA aims of helping countries to identify:

> "… areas of relative strength and weakness [and] monitor progress and raise aspirations"

and to:

> "… provide directions for national policy, for schools' curriculum and instructional efforts and for students' learning."

The KS3 components of reading performance or **Assessment Focuses (AFs)**, are as follows:

AF2 understand, describe, select or retrieve information, events or ideas from texts

AF3 deduce, infer or interpret information, events or ideas from texts

AF4 identify and comment on writers' uses of language, including grammatical and literary features at text level

AF5 explain and comment on writers' uses of language, including grammatical and literary features at word and sentence level

AF6 identify and comment on writers' purposes and viewpoints and the overall effect of the text on the reader.

There are in addition two AFs that are not specifically targeted at KS3:

AF1 use a range of strategies, including accurate decoding of text, to read for meaning

AF7 relate texts to their social, cultural and historical contexts and literary tradition.

AF1 is relevant for teacher assessment at KS1 and AF7 is targeted at GCSE

level and above.

There are clear similarities between these assessment focuses and the skills targeted in the PISA tests, although the KS3 criteria are slightly more specific and they refer explicitly to responding to texts at word, sentence and text levels.

Examples of KS3 test questions for each of the AFs are as follows:

AF2 understand, describe, select or retrieve information, events or ideas from texts

From the first sentence in paragraph 1, write down one word which suggests the writer admires Lulworth Cove. (2004)

According to paragraph 1, what two desires motivate the majority of treasure-hunters. (2003

AF3 deduce, infer or interpret information, events or ideas from texts

From paragraphs 1 and 2, give two different reasons why the bin men's job is demanding. (2004)

Give one detail from paragraph 1 which shows that the narrator and his companions are walking on high ground. (2003)

AF4 identify and comment on writers' uses of language, including grammatical and literary features at text level

The web page begins:

For the last five years, something huge, strange, almost magical, has been growing in a giant crater, deep in the South-West of England.

Explain two ways this sentence makes the reader want to read more.

Support each explanation with a quotation from this sentence. (2004)

Explain two ways paragraph 1 creates an atmosphere of mystery and suspense at the beginning of the text. (2003)

AF5 explain and comment on writers' uses of language, including grammatical and literary features at word and sentence level

In paragraph 2, the writer describes the line of people as moving relentlessly.

What does the phrase moving relentlessly suggest about the people? (2004)

In paragraphs 4 and 5, how does the writer's use of language suggest the different impressions he gets of the treasure? (2003)

AF6 identify and comment on writers' purposes and viewpoints and the overall effect of the text on the reader.

Suggest one reason why the writer included a** range **of examples to support her ideas in paragraph 2. (2003)

How does the writer try to make the reader feel some responsibility for the problem of waste disposal?

You should comment on the effect of:

- the bin men's comments about rubbish:

- the statistics used:

- references to other countries. (2004)

6.3.3 GCSE English Examinations

At GCSE level, all of the awarding bodies are required to develop specifications that target reading **Assessment Objectives (AOs)** as laid down by the Qualifications and Curriculum Authority (QCA):

Reading

Candidates must demonstrate their ability to:

(i) read with insight and engagement, making appropriate references to texts and developing and sustaining interpretations of them;

(ii) distinguish between fact and opinion and evaluate how information is presented;

(iii) follow an argument, identifying implications and recognising inconsistencies;

(iv) select material appropriate to their purpose, collate material from different sources and make cross-references;

(v) understand and evaluate how writers use linguistic, structural and presentational devices to achieve their effects and comment on ways language varies and changes.

QCA also specifies the range of texts that must be covered on GCSE English courses:

- *a play by Shakespeare*
- *work from the English literary heritage by at least one major writer with a well established critical reputation*
- *texts from different cultures and traditions*
- *non-fiction texts*
- *media texts.*

The following pages provide examples of representative GCSE questions at both Foundation and Higher Tiers. By far the most popular GCSE English specification is

AQA (Specification A) and examples are drawn from the June 2005 papers. Other examples are taken from the June 2005 OCR papers and an Edexcel example has already been shown. Some of these examples are similar to PISA items but it is important to recognise that although all GCSE specifications provide equivalent coverage of the reading curriculum, there are also significant differences between different awarding bodies' question paper structures and task-setting styles:

- texts representing the different reading categories may be studied in preparation for examination questions or they may be used as the basis for coursework assignments

- Foundation and Higher Tier examination questions may be based on the same texts, or slightly edited versions of the same texts in the case of the Foundation Tier, or entirely different texts may be used for the examination papers for each tier

- some awarding bodies use bullet points to provide a structure for students' responses on Foundation Tier papers but do not use them for Higher Tier papers

- other bodies use bullet points in similar ways on both tiers to help students to structure their responses

- some bodies use a mixture of low tariff and high tariff reading questions for a single text on Foundation Tier papers but set only one more broadly based question per text for Higher tier papers.

All of these approaches are illustrated in the examples below.

OCR June 2005

<u>English (Specification 1900): Unit 1 Non-fiction, Media and Information</u>

<u>Reading: NON-FICTION</u>

<u>Paper 1 (Foundation Tier)</u>

Stimulus texts: passage from *I Can't Stay Long* by Laurie Lee

travel brochure *Welcome to Amsterdam*

1 (a) In the first paragraph, what does the author say the Hollanders did to dry out their country? [2]

(b) From the second paragraph, give one **fact** about Schipol airport.

Give **two** words which show the author's **opinions** of the old houses by the canals.

[3]

(c) What does Laurie Lee find interesting about Holland?

In your answer, you should write about:

- Amsterdam
- The countryside
- The lives of the people in the city and the countryside.

Use **your own words** as far as possible.

[15]

2 Show how 'Welcome to Amsterdam' presents the city as an attractive place for tourists to visit.

In your answer you should comment on:

- The choice and position of the photograph
- The details you are given about the city
- The words and phrases used.

[20]

Reading: NON-FICTION

Paper 2 (Higher Tier)

Read both passages carefully. **Question 1** refers to *Disneyland: The Beginning* and question 2 refers to *Park and Pride*.

These answers will be marked for reading. Plan your answers and write them carefully.

1 By referring closely to the article *Disneyland: The Beginning*, outline concisely what the writer says about how Disneyland was planned and financed, the problems that arose when it was opened, and the reasons he gives as to why it was different from any previous amusement parks.

Use your own words as far as possible. [30]

Now read *Park and Pride* and answer question 2.

2 Explore Martin Palmer's attitude to his rollercoaster ride and how he attempts to make his readers share his experiences of the theme park. In your answer consider closely how he presents his experiences, and how the language he uses helps to communicate his feelings.

[30]

English (Specification 1900): Unit 2 Different Cultures, Analysis and Argument

Reading: Texts from Different Cultures and Traditions

Paper 1 (Foundation Tier)

[Candidates answer one question from a choice of two on the set text they have studied. The first question in each case is based on a short extract from the text whereas the second is not supported by an extract.]

Either

5 Remind yourself of the passage below, and then answer the question that follows.

[Okonkwo ruled…was idleness. (19 lines of text)]

What was it about his father that Okonkwo hated? How did this hatred affect his behaviour in one or two other episodes in the novel?

Support your answer by referring to and quoting from the novel.

Remember to put quotation marks round any words and phrases you use from the novel.

(The equivalent Higher Tier question is:

How does the writer show that feat dominates Okonkwo here and in the rest of the novel?)

Or

6 What are the differences between the three white men who bring change to Umuofia?

Support your answer by referring to and quoting from the novel.

Remember to put quotation marks round any words and phrases you use from the novel.

(The equivalent Higher Tier question is:

How does Achebe present the differing characters of the white men who bring change to Umuofia?)

AQA (Specification A) June 2005

Paper 1 (Foundation Tier)

1 Read **Item 1**, the advertisement called *Gadgets for the Girls*.

(a) Select and write down four points which support the argument

that this car would suit 'the girls'. *(4 marks)*

(b) Now read **Item 2** the extract called *Why No One Walks* from a book by Bill Bryson.

What does the extract show you:

- about American attitudes to walking
- about Bill Bryson's attitude towards walking? *(8 marks)*

Now look at **Item 1** and **Item 2** together.

(c) Compare the views about cars in Item 1 with the views about cars in Item 2 *(5 marks)*

2 Now look again at **Item 1**.

How effective is this advertisement in making the car seem attractive?

Write about the effectiveness of:

- the layout of the page
- the pictures
- the uses of language. *(10 marks)*

AQA (Specification A) June 2005

Paper 1 (Higher Tier)

1 Read **Item 1**, the extract called *Why No One Walks* from a book by Bill Bryson.

(a) What surprises Bryson about the way Americans live? *(3 marks)*

(b) What methods does Bryson use to entertain the reader? *(4 marks)*

Read **Item 2,** the advertisements called *Gadgets for Girls*.

(c) Compare the views about cars in Item 1 with the views about cars in Item 2 *(6 marks)*

2 Look again at **Item 2,** the advertisement called *Gadgets for Girls*.

(a) How does the use of language in the advertisement make the car seem desirable? *(8 marks)*

(b) How effective are the pictures in helping to support the claims made for the car in the written text? *(6 marks)*

AQA (Specification A) June 2005

Paper 2 (Foundation Tier)

Section A: Reading: Poetry from Different Cultures

Either

1 Compare the ways language and layout are used to achieve particular effects in *Two Scavengers in a Truck, Two Beautiful People in a Mercedes* (page 8) and **one** other poem of your choice from the Poems from Different Cultures.

Compare:

- the kinds of language and layout used
- the effects achieved by these uses of language and layout
- your own response to the poems. *(27 marks)*

Or

2 Compare the ways in which the poets show how important personal identity is in *Half-Caste* (page 13) and **one** other poem from Poems

from Different Cultures.

Compare:

- the feelings about identity in the poems
- how the poets show the importance of being who you are
- how the poets use language to show ideas and feelings about who you are and how important that is. *(27 marks)*

AQA (Specification A) June 2005

Paper 2 (Higher Tier)

Section A: Reading: Poetry from Different Cultures

Either

1 How is contrast used in *Two Scavengers in a Truck, Two Beautiful People in a Mercedes* (page 8)?

Compare this with the use of contrast in one other poem. *(27 marks)*

Or

2 Compare how identity is shown to be important in *Half-Caste* (page

13) and **one** other poem. *(27 marks)*

6.3.4 PISA, KS3 English and GCSE English Compared

In summary, the comparison of PISA reading literacy and the two relevant assessments in England showed a range of similarities and some differences. The answers to the questions raised at the beginning of this section are:

How far does PISA reading reflect what we call school subject English?

1 The assessment criteria underpinning PISA, KS3 tests and GCSE reading literacy items are largely compatible. Differences of nomenclature in the respective assessment focuses or objectives are mainly cosmetic.

2 The balance of reading genre covered by the PISA units differs from that of the KS3 and KS4 reading curriculum. By comparison, literary forms are under-represented in the PISA units and non-fiction and transactional texts are over-represented.

3 PISA reading units contain non-continuous texts of a kind that would not normally be considered part of the KS3 and KS4 reading curriculum.

On what basis can we say how familiar and accessible is PISA reading?

4 Familiarity and accessibility are functions of:

a) the extent to which the linguistic demands of a text are appropriate to the reading abilities of the target readership

b) the extent to which the target audience is interested in the subject matter presented in the texts

c) the quality of presentation, including the choices of font, point size, line length, the use of columns and other design features which may enhance the readability of the page.

Some of the PISA texts, by these criteria, are less familiar and less accessible than is typically the case with KS3 and GCSE texts.

5 KS3 and PISA low-tariff items aimed at testing information retrieval are often similar in approach.

6 GCSE reading questions often target interpretation and evaluation, and leave students free to select or retrieve content points relevant to their purposes, whereas PISA items tend to be more specific in targeting precise parts of the text.

7 All PISA units are necessarily encountered as unseen by pupils, as is the case at KS3 (apart from the Shakespeare paper). At GCSE, depending on the structure of individual specifications, students encounter some texts as unseen whereas others are studied in advance of the examination. This clearly affects the nature of item setting and expectations about what students may be expected to achieve in a given time.

How far are pupils in England called upon to offer an opinion in tests and exams?

8 Pupils' formulation and expression of opinions are central to the English curriculum in England even during the pre-KS3 phase. The reading assessment focuses 3, 4, 5 and 6 are all relevant to KS3 test approaches to evaluation, interpretation and reflection about texts. At GCSE level, Reading AOs (i), (ii), (iii) and (v) are all relevant to these aspects of reading.

7. Conclusions and Discussion

The study comprised two main strands, a rating exercise on the familiarity of TIMSS and PISA items and a comparison of TIMSS, PISA, KS3 tests and GCSE papers. Mathematics and science could be examined in all four assessments, although in PISA these are called mathematical literacy and scientific literacy, reflecting PISA's approach. The familiarity of the PISA reading literacy items was also rated, and the PISA assessment compared with KS3 and GCSE, there being no equivalent IEA reading comprehension study in the secondary school years.

TIMSS and PISA Compared

The rating study looked at each item separately. In mathematics and science the ratings were of the familiarity of the concept assessed, the context in which it was set and the format and presentation of the item. A general rating for appropriateness to students in England was also made.

- In science the results showed no difference between PISA and TIMSS items in terms of either familiarity or appropriateness.
- In mathematics, however, the TIMSS items were rated as more appropriate overall, and the TIMSS contexts and item formats were found to be more familiar.
- Such differences between PISA and TIMSS were not found in the familiarity of the concept being assessed, ie the mathematics itself.

TIMSS, PISA and National Tests Compared

In the comparison of the TIMSS, PISA, KS3 and GCSE questions and papers the assessments were considered on a broader scale than in the rating exercise.

- It was found that all four assessments drew from the same pool of item types, with multiple choice, short response and extended response items all being used, although the proportions of the different types varied.

- Based on KS3 or GCSE practice, pupils in England are familiar with each of these types.

- The fact that TIMSS employs a high proportion of multiple choice items would not, therefore, be expected to either advantage or disadvantage students in England on the grounds of familiarity.

- PISA mathematics contains a greater proportion than TIMSS of longer tasks, in which a large scale context is used to derive several marks, but both also contain a wide range of shorter questions.

- The KS3 mathematics tests and GCSE also include questions in a range of contexts, and short items out of context. Again no advantage or disadvantage on grounds of familiarity is to be expected when students are faced with TIMSS or PISA.

- TIMSS uses the same range of item types in roughly the same proportions for both mathematics and science.

- PISA science, as judged by the limited number of questions used in 2000 and 2003, consistently adopts the large scale context with several marks. This, however, is also characteristic of KS3 science tests and, to a lesser extent, GCSE science examinations. Again, there is no reason, on familiarity grounds, to expect this to either favour or disadvantage students in England.

In summary, the situation is that the ingredients of TIMSS, PISA, KS3 and GCSE are the same. None the less, the recipes used do differ, so the quantities of the ingredients are not the same. The differences, however, involve varying proportions of styles of assessment and variations in the use of context that are within parameters that students in England are already familiar with before taking a TIMSS or PISA test.

- One significant difference between PISA and the other assessments is in the amount of reading required. This is apparent in all of the PISA science questions and in some of those in mathematics, and is not something that students in England would be familiar with.

We have considered changes to the various assessments involved since the late 1990s. In the national context the changes at KS3 have most pertinently concerned the increased assessment of attainment target 1. This has introduced new types of item into KS3 tests, but has not affected the item types used in ways relevant to comparisons with international surveys. Similarly changes within GCSE have not changed the comparative situation. Again, the incorporation of longer problem solving and inquiry tasks into TIMSS in 2003 has not involved the introduction of anything unfamiliar to students in England.

The situation is complex, and one issue remains to be considered: could differences in familiarity account for differential in performance in England between PISA and TIMSS?

- In general what students are asked to do in either PISA or TIMSS, particularly with regard to item type, is likely to be reasonably familiar to them so unfamiliarity per se is an unlikely causal factor in any differential performance between the two surveys.

- Proportions of item types do vary, however, and it is possible that, say, more multiple choice in TIMSS than in PISA might lead to differences in performance in spite of their familiarity to students in England.

- There is no evidence from this study that the heavier contextualisation in PISA is more familiar to students in England than other approaches; it lies within the range with which they are familiar.

- The quantity of reading required in many PISA questions does make them unfamiliar to students in England, and this remains, in our view, the most significant difference between PISA and TIMSS.

It is the quantity of reading that marks PISA out, not the complexity of the language, which is similarly unfamiliar in both the international studies. The high reading demand of questions in PISA is often accompanied by a relatively lower demand in the mathematics or science required. This reflects the lower level of mathematics or science that students can apply in new contexts as opposed to very familiar ones. The combined impact on performance in England is not clear cut. This illustrates a further

salient point. The study has focussed on familiarity, but the relationship between this and difficulty is likely to be complex. To investigate further it might be fruitful to build on the work on familiarity by looking for relationships between familiarity in content and item type and difficulty level. This would not be simple, as content and item type are not independent variables and the confounding effects of this might prove very difficult to disentangle within the item pool available for analysis.

Appendices

Samples of items from the PISA and TIMSS are given in appendices 1 to 5.

Appendix 1: Sample TIMSS Science Items

Example Science Items: **Grade 8**

nfer

1

The diagram above shows the Pacific Ring of Fire. Earthquakes and volcanic activity occur along the Ring of Fire. Which of the following best explains why?

* Ⓐ It is located at the boundaries of tectonic plates.

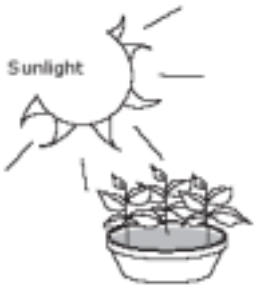
Ⓑ It is located at the boundary of deep and shallow water.

Ⓒ It is located where the major ocean currents meet.

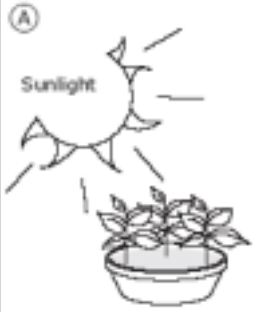
Ⓓ It is located where ocean temperature is the highest.

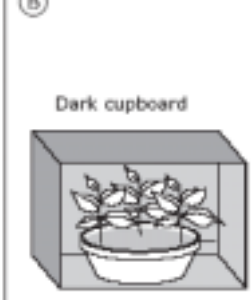
*Correct Answer

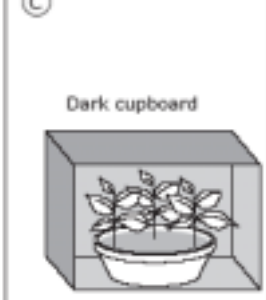
2 A girl has an idea that green plants need sand in the soil for healthy growth. In order to test her idea she uses two pots of plants. She sets up one pot of plants as shown below.

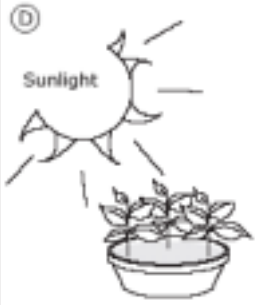


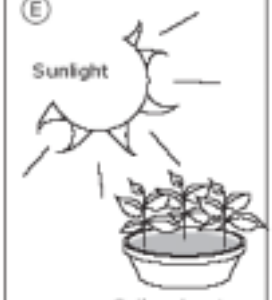
Which ONE of the following should she use for the second pot of plants?

Ⓐ


Ⓑ


Ⓒ


Ⓓ


★ Ⓔ


*Correct Answer

3

The picture shows how a student set up some apparatus in a laboratory for an investigation. The inverted test tube was completely filled with water at the beginning of the investigation as shown in Figure 1. After several hours, the level of water in the test tube had gone down as shown in Figure 2.

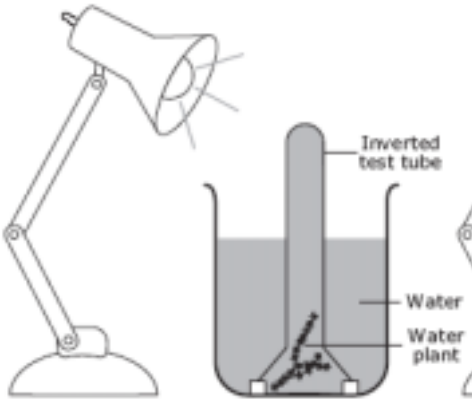


Figure 1

Figure 2

What is contained in the top part of the test tube labeled X in Figure 2?

(Check one box.)

- [ ] air
- [x] oxygen
- [ ] carbon dioxide
- [ ] vacuum

Explain your answer.

During photosynthesis, plants produce oxygen and glucose.

4

Three identical candles are placed in the three jars shown above and lit at the same time. Jars Y and Z are then sealed with lids, and Jar X is left open.

Which candle flame will go out first (X, Y, or Z)? Z

Explain your answer.

Z because fire needs oxygen to stay lit. With the lid being sealed no oxygen can get in. There is a little bit of air in there for it to stay lit. Since Z is smaller than Y, Z would go out first.

5

The diagram above shows a compass needle with its North and South poles labeled (N and S). It is placed next to a strong magnet as shown in the diagram below.

N S

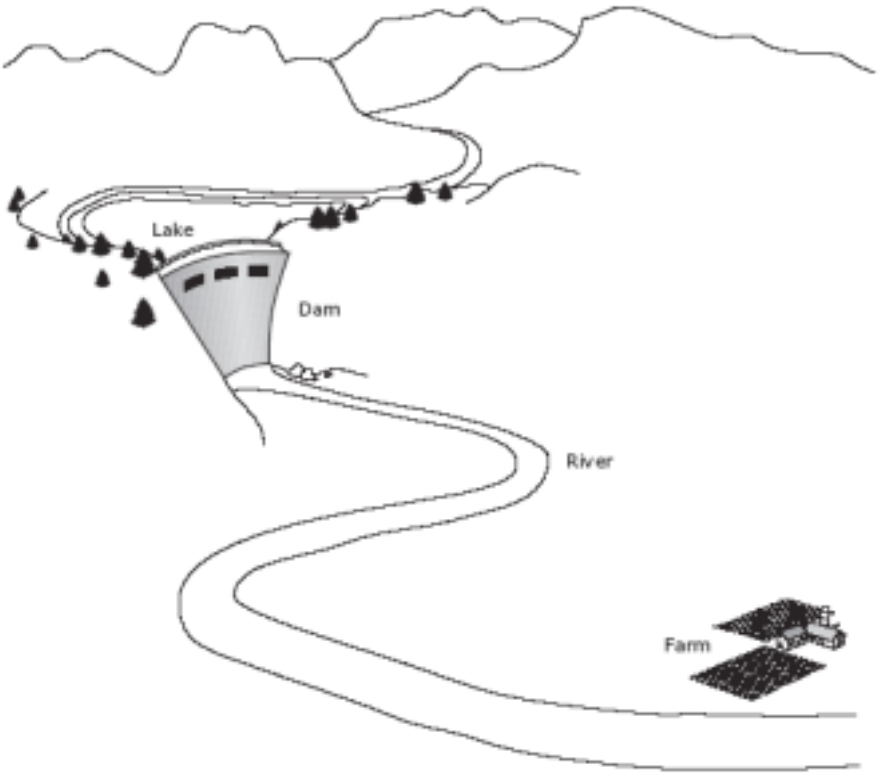
A. Draw the compass needle in the circle on the diagram above. Label the North (N) and South (S) poles of the needle.

B. Explain your answer using your knowledge of magnets.

The north pole on the magnet will attract the south pole. on the compass

6

The diagram shows a farm in a valley where a dam has just been built.



The presence of the dam can have both positive and negative effects on farming in the valley.

A. Describe one positive effect of the dam on farming.

The dam will stop the water from flooding the field.

B. Describe one negative effect of the dam on farming.

Less irrigation because the dam may not let enough water flow through.

Metal Crown

Instructions: Questions **7**, **8**, **9**, **10** are about Metal Crown. To answer these questions you may refer to any information shown on the pages in the Metal Crown section.

A king gave a jeweler a block of pure metal. He asked the jeweler to make him a crown out of the metal.

metal crown

metal block

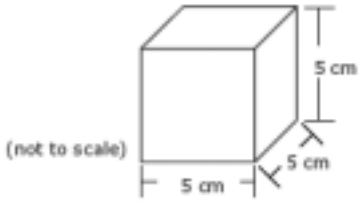
After the jeweler delivered the crown, the king observed it carefully. He thought that the jeweler might have substituted another pure metal or a mixture of metals to make the crown. He weighed the crown, and it had the same mass as the original block, 2400 grams. Still not satisfied, the king asked some scientists to help him find out what the crown was made of.

Questions for Metal Crown begin on the next page.

7

The scientists decided to compare the densities of the crown and a block of metal just like the originalblock. The density of a substance is the mass of a sample of the substance divided by its volume (density = mass/volume).

The scientists found the volume of the block and computed its density based on its known mass (2400g). The diagram below shows the dimensions of the block of metal that the scientists measured.



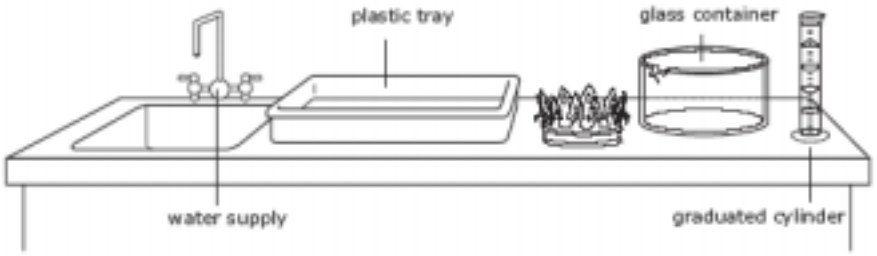
What is the density of the block of metal?

Answer: 19.2 g/cm^3

Questions for Metal Crown continue.

8

The scientists then needed to find the volume of the crown in order to determine its density. The following equipment and materials were available for them to use.



Describe a procedure that the scientists could use to find the volume of the crown using some or all of the equipment and materials shown above. You may use diagrams to help explain your procedure.

Fill the beaker with enough water to cover the crown. Add the crown and mark the side of the beaker where the water level is. Then take the crown out. Use the graduated cylinder to add little bits of water until the level comes back up to the mark. That is the volume of the crown.

Questions for Metal Crown continue.

9 The scientists measured the volume of the crown five times. They computed the density for each volume measurement. Their results are shown in the table below.

Trial	Volume of Crown (cm^3)	Density of Crown (g/cm^3)
1	202	11.88
2	200	12.00
3	201	11.94
4	198	12.12
5	199	12.06

A. Why did the scientists measure the volume five times?

Because there is experimental error. So, measuring it 5 times you can calculate the average to know how much error there is.

B. The scientists reported to the king that the density of the crown was 12.0 g/cm^3. Show how the scientists used their results to obtain this value for the density.

They added together all of the densities and then divided by 5 to get the average.

Questions for Metal Crown continue. ➡

10

The table below lists the density for different metals.

Metal	Density (g/cm³)
Platinum	21.4
Gold	19.3
Silver	10.5
Copper	8.9
Zinc	7.1
Aluminum	2.7

A. Look at the density you computed for the block of metal. What was the block of metal most likely made of?

Answer: Gold

Explain your answer.

It had the closest density.

B. The density of the crown was found to be 12.0 g/cm³. What would you report to the king about what metal or mixture of metals the jeweler used to make the crown?

The jeweler used some silver as well as gold.

End of Metal Crown section.

Appendix 2: Sample TIMSS Mathematics Items

Example Mathematics Items: Grade 8

1

If n is a negative integer, which of these is the largest number?

Ⓐ $3 + n$

Ⓑ $3 \times n$

*Ⓒ $3 - n$

Ⓓ $3 \div n$

2

In a car rally two checkpoints are 160 km apart. Drivers must travel from one checkpoint to the other in exactly 2.5 hours to earn maximum points.

A. What must the average speed be to travel the 160 km in this time?

Answer: 64 kph

$2.5 \overline{)160}$ 64

14
3.5
12
3

2.5
5
7.5
10

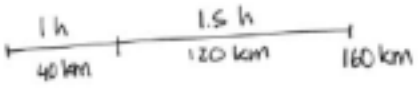
B. A driver took 1 hour to travel through a 40 km hilly section at the beginning of the course.

What must the average speed, in kilometers per hour, be for the remaining 120 km if the total time between checkpoints is to be 2.5 hours?

Answer: 80 kph

$1.5 \overline{)120}$ 80

8
4
12



*Correct Answer

3

The three figures below are divided into small congruent triangles.

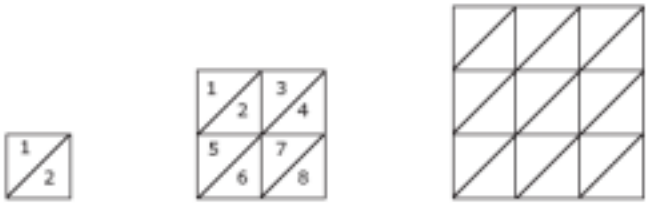


Figure 1 Figure 2 Figure 3

A. Complete the table below. First, fill in how many small triangles make up Figure 3. Then, find the number of small triangles that would be needed for the 4th figure if the sequence of figures is extended.

Figure	Number of Small Triangles
1	2
2	8
3	18
4	32

B. The sequence of figures is extended to the 7th figure. How many small triangles would be needed for Figure 7?

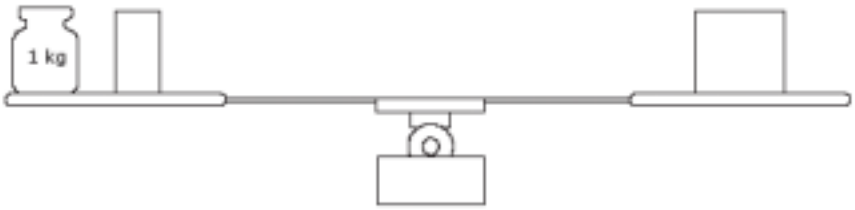
Answer: 98

C. The sequence of figures is extended to the 50th figure. Explain a way to find the number of small triangles in the 50th figure that does not involve drawing it and counting the number of triangles.

multiply the figure by itself and then multiply the answer you get by two.

4

The objects on the scale make it balance exactly. On the left pan there is a 1 kg weight (mass) and half a brick. On the right pan there is one brick.



What is the weight (mass) of one brick?

Ⓐ 0.5 kg

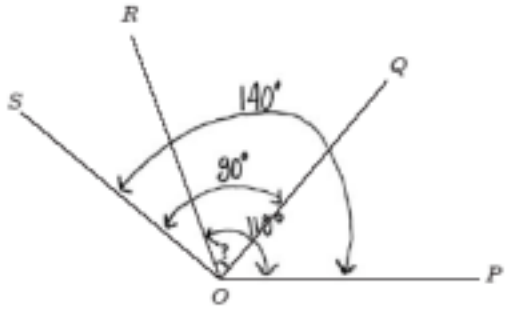
Ⓑ 1 kg

* Ⓒ 2 kg

Ⓓ 3 kg

5

In the figure, the measure of $\angle POR$ is 110°, the measure of $\angle QOS$ is 90°, and the measure of $\angle POS$ is 140°.



What is the measure of $\angle QOR$?

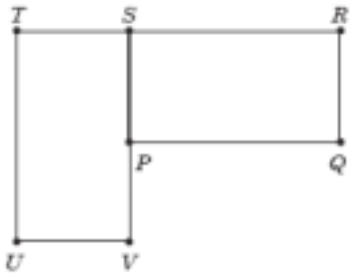
$140 - 110 = 30$
$90 - 30 = 60$

Answer: $60°$

*Correct Answer

6

Rectangle $PQRS$ can be rotated (turned) onto rectangle $UVST$.



What point is the center of rotation?

Ⓐ P

Ⓑ R

* Ⓒ S

Ⓓ T

Ⓔ V

*Correct Answer

Phone Plans

Instructions: Questions **7, 8, 9** are about Phone Plans.
To answer these questions you may refer to any information shown on the pages in the Phone Plans section.

Betty, Frank, and Darlene have just moved to Zedland. They each need to get phone service. They received the following information from the telephone company about the two different phone plans it offers.

They must pay a set fee each month and there are different rates for each minute they talk. These rates depend on the time of the day or night they use the phone, and on which payment plan they choose. Both plans include time for which phone calls are free. Details of the two plans are shown in the table below.

Plan	Monthly Fee	Rate per minute		Free minutes per month
		Day (8 am – 6 pm)	Night (6 pm – 8 am)	
Plan A	20 zeds	3 zeds	1 zed	180
Plan B	15 zeds	2 zeds	2 zeds	120

7 Betty talks for less than 2 hours per month. Which plan would be less expensive for her?

Less expensive plan B

Explain your answer in terms of both the monthly fee and free minutes.

2 hours = 120 min. Only uses the free minutes

A: 20 + 0 = 20

B: 15 + 0 = 15 Cheaper

Questions for Phone Plans continue.

8 Frank talks for 5 hours per month at the night rate. What would each plan cost him per month? Show your work.

Cost Per Month for Plan A: 140 zeds

Cost Per Month for Plan B: 375 zeds

5 hrs = 5 × 60 = 300 min

A: 300 − 180 = 120 min
120 × 1 = 120 zeds
120 + 20 = 140 zeds.

B: 300 − 120 = 180 min
180 × 2 = 360 zeds.
360 + 15 = 375 zeds.

Questions for Phone Plans continue.

9 Darlene signed up for the *Plan B*, and the cost of one month of service was 75 zeds. How many minutes did she talk that month? Show your work.

Minutes talked 150

75 − 15 = 60 zeds
60 : 2 = 30 minutes
30 + 120 = 150 minutes

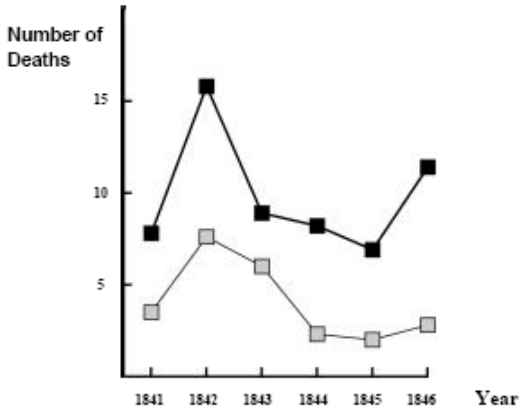
End of Phone Plans section.

Appendix 3: Sample PISA Science Items

SEMMELWEIS' DIARY TEXT 1

"*July 1846. Next week I will take up a position as 'Herr Doktor' at the First Ward of the maternity clinic of the Vienna General Hospital. I was frightened when I heard about the percentage of patients who die in this clinic. This month not less than 36 of the 208 mothers died there, all from puerperal fever. Giving birth to a child is as dangerous as first-degree pneumonia.*"

These lines from the diary of Ignaz Semmelweis (1818-1865) illustrate the devastating effects of puerperal fever, a contagious disease that killed many women after childbirth. Semmelweis collected data about the number of deaths from puerperal fever in both the First and the Second Wards (see diagram).



Physicians, among them Semmelweis, were completely in the dark about the cause of puerperal fever. Semmelweis' diary again:

"December 1846. Why do so many women die from this fever after giving birth without any problems? For centuries science has told us that it is an invisible epidemic that kills mothers. Causes may be changes in the air or some extraterrestrial influence or a movement of the earth itself, an earthquake."

Nowadays not many people would consider extraterrestrial influence or an earthquake as possible causes of fever. But in the time Semmelweis lived, many people, even scientists, did! We now know it has to do with hygienic conditions. Semmelweis knew that it was unlikely that fever could be caused by extraterrestrial influence or an earthquake. He pointed at the data he collected (see diagram) and used this to try to persuade his colleagues.

Question 1: SEMMELWEIS' DIARY

S195Q02- 01 02 03 04 11 12 13 21 9

Suppose you were Semmelweis. Give a reason (based on the data Semmelweis collected) why puerperal fever is unlikely to be caused by earthquakes.

.....

.....

.....

.....

.....

SEMMELWEIS' DIARY TEXT 2

Part of the research in the hospital was dissection. The body of a deceased person was cut open to find the cause of death. Semmelweis recorded that the students working on the First ward usually took part in dissections on women who died the previous day, before they examined women who had just given birth. They did not pay much attention to cleaning themselves after the dissections. Some were even proud of the fact that you could tell by their smell that they had been working in the mortuary, as this showed how industrious they were!

One of Semmelweis' friends died after having cut himself during such a dissection. Dissection of his body showed he had the same symptoms as mothers who died from puerperal fever. This gave Semmelweis a new idea.

Question 2: SEMMELWEIS' DIARY

S195Q

Semmelweis' new idea had to do with the high percentage of women dying in the maternity wards and the students' behaviour.

What was this idea?

A Having students clean themselves after dissections should lead to a decrease in puerperal fever.
B Students should not take part in dissections because they may cut themselves.
C Students smell because they do not clean themselves after a dissection.
D Students want to show that they are industrious, which makes them careless when they examine the women.

Question 3: SEMMELWEIS' DIARY

S195Q05- 01 02 11 12 13 14 15

Semmelweis succeeded in his attempts to reduce the number of deaths due to puerperal fever. But puerperal fever even today remains a disease that is difficult to eliminate.

Fevers that are difficult to cure are still a problem in hospitals. Many routine measures serve to control this problem. Among these measures is washing sheets at high temperatures.

Explain why high temperature (while washing sheets) helps to reduce the risk that patients will contract a fever.

.....

.....

Question 4: SEMMELWEIS' DIARY

S195Q

Many diseases may be cured by using antibiotics. However, the success of some antibiotics against puerperal fever has diminished in recent years.

What is the reason for this?

A Once produced, antibiotics gradually lose their activity.
B Bacteria become resistant to antibiotics.
C These antibiotics only help against puerperal fever, but not against other diseases.
D The need for these antibiotics has been reduced because public health conditions have improved considerably in recent years.

OZONE TEXT

Read the following section of an article about the ozone layer.

The atmosphere is an ocean of air and a precious natural resource for sustaining life on Earth. Unfortunately, human activities based on national/personal interests are causing harm to this common resource, notably by depleting the fragile ozone layer, which acts as a protective shield for life on Earth.

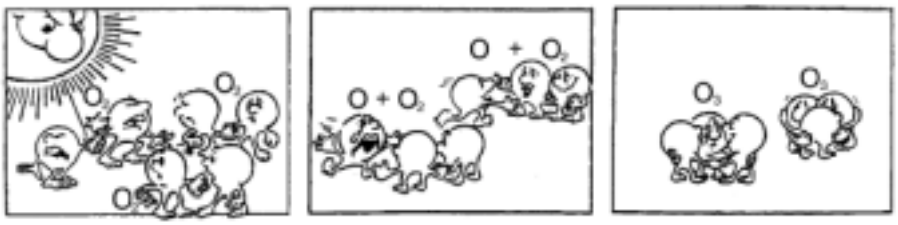
Ozone molecules consist of three oxygen atoms, as opposed to oxygen molecules which consist of two oxygen atoms. Ozone molecules are exceedingly rare: fewer than ten in every million molecules of air. However, for nearly a billion years, their presence in the atmosphere has played a vital role in safeguarding life on Earth. Depending on where it is located, ozone can either protect or harm life on Earth. The ozone in the troposphere (up to 10 kilometres above the Earth's surface) is 'bad' ozone which can damage lung tissues and plants. But about 90 percent of ozone found in the stratosphere (between 10 and 40 kilometres above the Earth's surface) is 'good' ozone which plays a beneficial role by absorbing dangerous ultraviolet (UV-B) radiation from the Sun.

Without this beneficial ozone layer, humans would be more susceptible to certain diseases due to the increased incidence of ultra-violet rays from the Sun. In the last few decades the amount of ozone has decreased. In 1974 it was hypothesised that chlorofluorocarbons (CFCs) could be a cause for this. Until 1987, scientific assessment of the cause and effect relationship was not convincing enough to implicate CFCs. However, in September 1987, diplomats from around the world met in Montreal (Canada) and agreed to set sharp limits to the use of CFCs.

Question 5: OZONE

S253Q01- 01 11 12 13 21 22 23 31 99

In the text above nothing is mentioned about the way ozone is formed in the atmosphere. In fact each day some ozone is formed and some other ozone disappears. The way ozone is formed is illustrated in the following comic strip.



Suppose you have an uncle who tries to understand the meaning of this strip. However, he did not get any science education at school and he doesn't understand what the author of the strip is explaining. He knows that there are no little fellows in the atmosphere but he wonders what those little fellows in the strip stand for, what those strange notations O, O_2 and O_3 mean and which processes the strip represents. He asks you to explain the strip. Assume that your uncle knows:

- that O is the symbol for oxygen;
- what atoms and molecules are.

Write an explanation of the comic strip for your uncle.

In your explanation, use the words atoms and molecules in the way they are used in lines 5 and 6.

.....

.....

.....

.....

.....

.....

Question 6: OZONE

S253Q02

Ozone is also formed during thunderstorms. It causes the typical smell after such a storm. In lines 9–13 the author of the text distinguishes between 'bad ozone' and 'good ozone'.

In terms of the article, is the ozone that is formed during thunderstorms 'bad ozone' or 'good ozone'?

Choose the answer and the explanation that is supported by the text.

	Bad ozone or good ozone?	Explanation
A	Bad	It is formed during bad weather.
B	Bad	It is formed in the troposphere.
C	Good	It is formed in the stratosphere.
D	Good	It smells good.

Question 7: OZONE

S253Q05-0 1 9

Lines 14 and 15 state: "Without this beneficial ozone layer, humans would be more susceptible to certain diseases due to the increased incidence of ultra-violet rays from the sun."

Name one of these specific diseases.

.....

Question 8: OZONE

S253Q03

At the end of the text, an international meeting in Montreal is mentioned. At that meeting lots of questions in relation to the possible depletion of the ozone layer were discussed. Two of those questions are given in the table below.

Which of the questions below can be answered by scientific research?

Circle Yes or No for each.

Question:	Answerable by scientific research?
Should the scientific uncertainties about the influence of CFCs on the ozone layer be a reason for governments to take no action?	Yes / No
What would the concentration of CFCs be in the atmosphere in the year 2002 if the release of CFCs into the atmosphere takes place at the same rate as it does now?	Yes / No

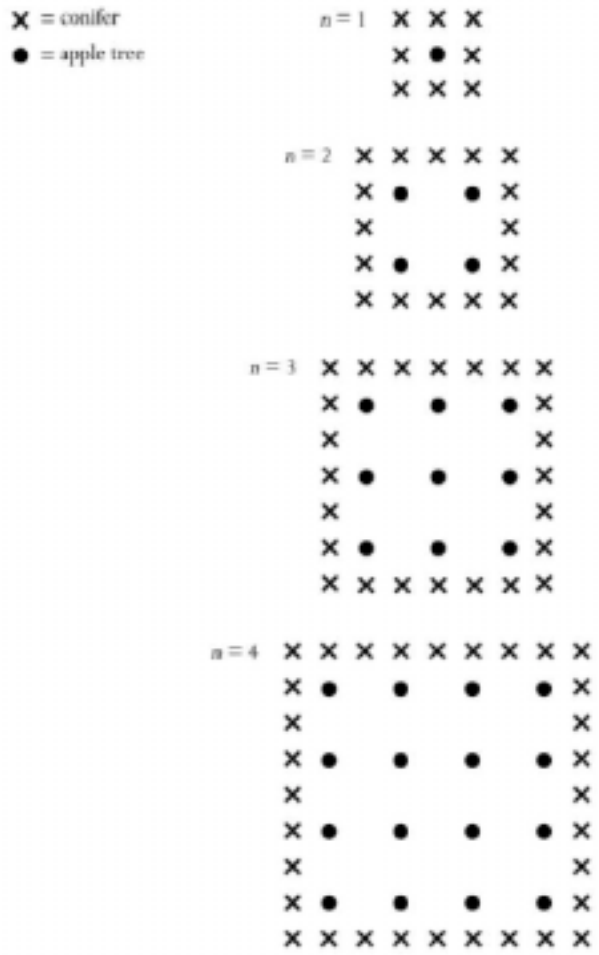
Appendix 4: Sample PISA Mathematics Items

MATHEMATICS UNIT 1

Apples

A farmer plants apple trees in a square pattern. In order to protect the trees against the wind he plants conifers all around the orchard.

Here you see a diagram of this situation where you can see the pattern of apple trees and conifers for any number (n) of rows of apple trees:



Question 1: APPLES

Complete the table:

n	Number of apple trees	Number of conifers
1	1	8
2	4	
3		
4		
5		

Question 2: APPLES

There are two formulae you can use to calculate the number of apple trees and the number of conifers for the pattern described above:

Number of apple trees = n^2

Number of conifers = $8n$

where n is the number of rows of apple trees.

There is a value of n for which the number of apple trees equals the number of conifers. Find the value of n and show your method of calculating this.

Question 3: APPLES

Suppose the farmer wants to make a much larger orchard with many rows of trees. As the farmer makes the orchard bigger, which will increase more quickly: the number of apple trees or the number of conifers?

Explain how you found your answer.

Continent area

Here is a map of Antarctica.

Question 1: CONTINENT AREA

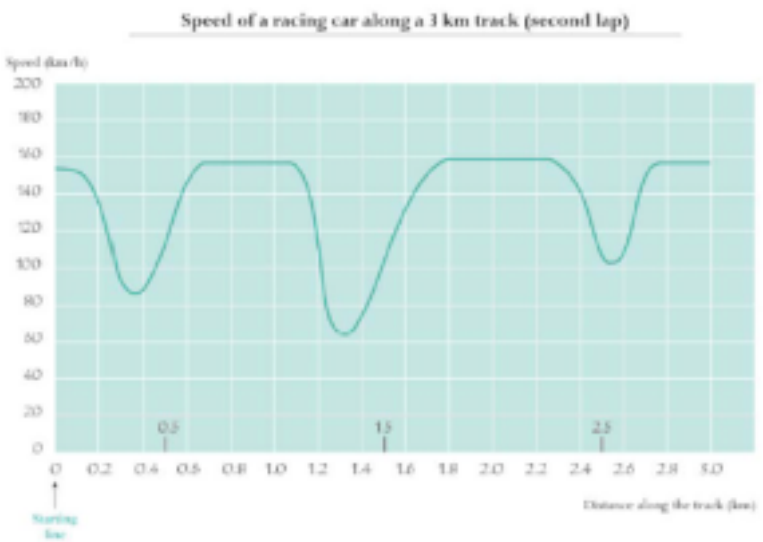
Estimate the area of Antarctica using the map scale.

Show your working out and explain how you made your estimate.
(You can draw over the map if it helps you with your estimation)

MATHEMATICS UNIT 3

Speed of a racing car

This graph shows how the speed of a racing car varies along a flat 3 kilometer track during its second lap.



Source: In memory of Claude Janvier, who died in June 1998. Modified task after his ideas in Janvier, C. (1978): The interpretation of complex graphs – studies and teaching experiments. Accompanying brochure to the Dissertation. University of Nottingham, Shell Centre for Mathematical Education, Item C-2.

The pictures of the tracks are taken from Fischer, R. & Malle, G. (1985): *Mensch und Mathematik*. Bibliographisches Institut: Mannheim-Wien-Zurich, 234-238.

Question 1: SPEED OF A RACING CAR

What is the approximate distance from the starting line to the beginning of the longest straight section of the track?

A 0.5 km.

B 1.5 km.

C 2.3 km.

D 2.6 km.

Question 2: SPEED OF A RACING CAR

Where was the lowest speed recorded during the second lap?

A At the starting line.

B At about 0.8 km.

C At about 1.3 km.

D Halfway around the track.

Question 3: SPEED OF A RACING CAR

What can you say about the speed of the car between the 2.6 km and 2.8 km marks?

A The speed of the car remains constant.

B The speed of the car is increasing.

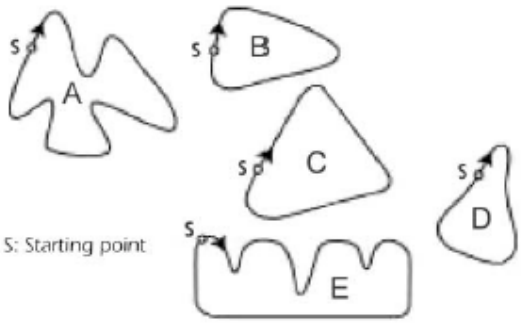
C The speed of the car is decreasing.

D The speed of the car cannot be determined from the graph.

Question 4: SPEED OF A RACING CAR

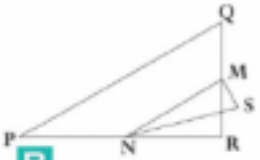
Here are pictures of five tracks:

Along which one of these tracks was the car driven to produce the speed graph shown earlier?

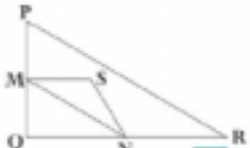


Triangles

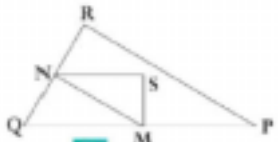
A



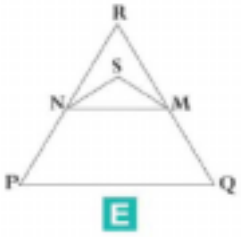
B



C



D



E

Question 1: TRIANGLES

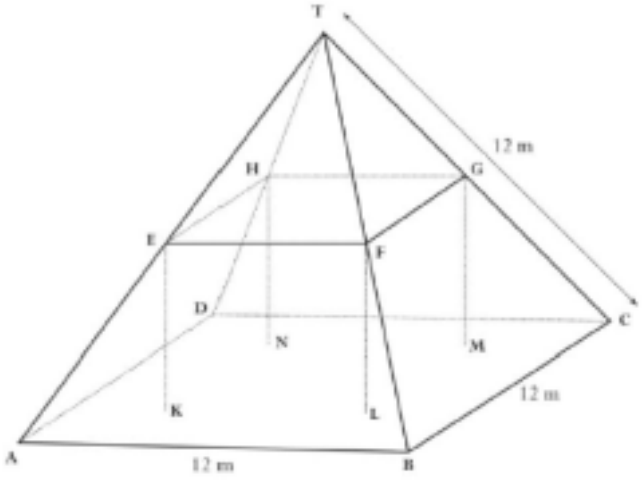
Circle the one figure on the previous page that fits the following description.

Triangle PQR is a right triangle with right angle at R. The line RQ is less than the line PR. M is the midpoint of the line PQ and N is the midpoint of the line QR. S is a point inside the triangle. The line MN is greater than the line MS.

Farms

Here you see a photograph of a farmhouse with a roof in the shape of a pyramid.

Below is a student's mathematical model of the farmhouse **roof** with measurements added.



The attic floor, ABCD in the model, is a square. The beams that support the roof are the edges of a block (rectangular prism) EFGHKLMN. E is the middle of AT, F is the middle of BT, G is the middle of CT and H is the middle of DT. All the edges of the pyramid in the model have the length 12 m.

Question 1: FARMS

Calculate the area of the attic floor ABCD.

The area of the attic floor ABCD = _____ m^2

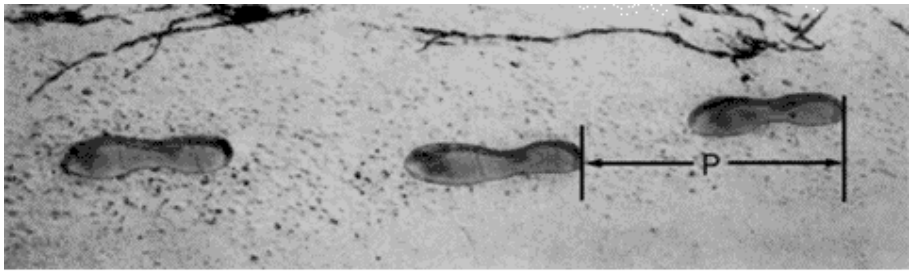
Question 2: FARMS

Calculate the length of EF, one of the horizontal edges of the block.

The length of EF = _____ m

PISA 2003 Released Maths Items

Walking



The picture shows the footprints of a man walking. The pacelength P is the distance between the rear of two consecutive footprints.

For men, the formula, $\frac{n}{P} = 140$, gives an approximate relationship between n and P where,

n = number of steps per minute, and

P = pacelength in meters.

Walking - Question #1

Question Sample Question

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

Answer Sample Answer

The correct answer is :
0.5 m or 50 cm, ½ (unit not required).

- 70/ p = 140
  70 = 140 p
  p = 0.5.
- 70/140.

Walking - Question #2

Question Sample Question

Bernard knows his pacelength is 0.80 meters. The formula applies to Bernard's walking.

Calculate Bernard's walking speed in meters per minute and in kilometers per hour. Show your work.

Answer Sample Answer

Correct answers (unit not required) for both metres/minute and km/hour:
n = 140 x .80 = 112.
Per minute he walks 112 x .80 metres = 89.6 metres.

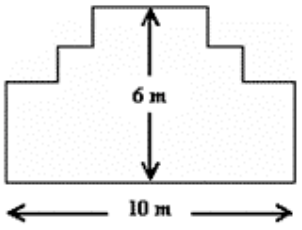
His speed is 89.6 metres per minute.
So his speed is 5.38 or 5.4 km/hr.

As long as both correct answers are given (89.6 and 5.4), whether working out is shown or not. Note that errors due to rounding are acceptable. For example, 90 metres per minute and 5.3 km/hr (89 X 60) are acceptable.

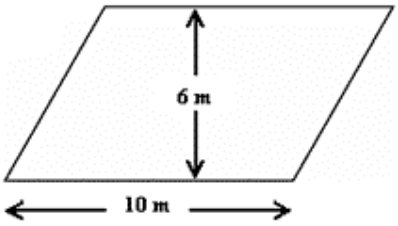
- 89.6, 5.4.
- 90, 5.376 km/h.
- 89.8, 5376 m/hour [note that if the second answer is given without units, it should be coded as 22] .

Students who show evidence of using a correct method, but provided an incorrect answer, received partial credit.

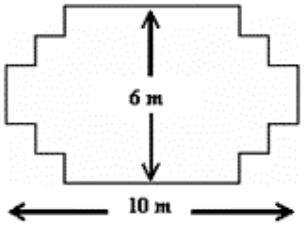
Carpenter



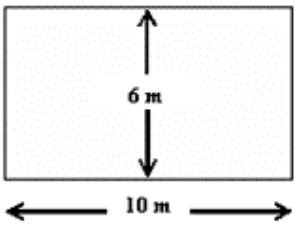
Design A



Design B



Design C



Design D

Garden bed design	Using this design, can the garden bed be made with 32 meters of timber?
Design A	Yes / No
Design B	Yes / No
Design C	Yes / No
Design D	Yes / No

A carpenter has 32 meters of timber and wants to make a border around a garden bed. He is considering the above designs for the garden bed.

Carpenter - Question #1

Question Sample Question

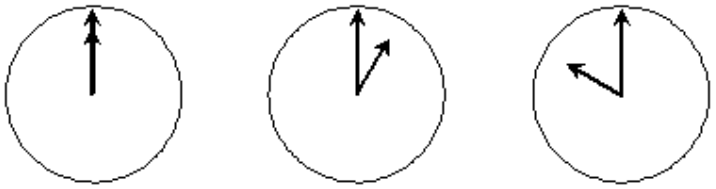
Circle either “Yes” or “No” for each design to indicate whether the garden bed can be made with 32 meters of timber.

Answer Sample Answer

The correct answer is :

Design A - Yes
Design B - No
Design C - Yes
Design D - Yes

Internet Relay Chat



Place	Time
Sydney	
Berlin	

Mark (from Sydney, Australia) and Hans (from Berlin, Germany) often communicate with each other using “chat” on the Internet. They have to log on to the Internet at the same time to be able to chat.

To find a suitable time to chat, Mark looked up a chart of world times and found the above:

Internet Relay Chat - Question #1

Question Sample Question

At 7:00 PM in Sydney, what time is it in Berlin?

Answer Sample Answer

The correct answer is:

10 AM or 10:00.

Internet Relay Chat - Question #2

Question Sample Question

Mark and Hans are not able to chat between 9:00 AM and 4:30 PM their local time, as they have to go to school. Also, from 11:00 PM till 7:00 AM their local time they won’t be able to chat because they will be sleeping.

When would be a good time for Mark and Hans to chat? Write the local times in the table.

Answer Sample Answer

The correct answer is:

Any time or interval of time satisfying the 9 hours time difference and taken from one of these intervals:
Sydney: 4:30 PM – 6:00 PM; Berlin: 7:30 AM – 9:00 AM

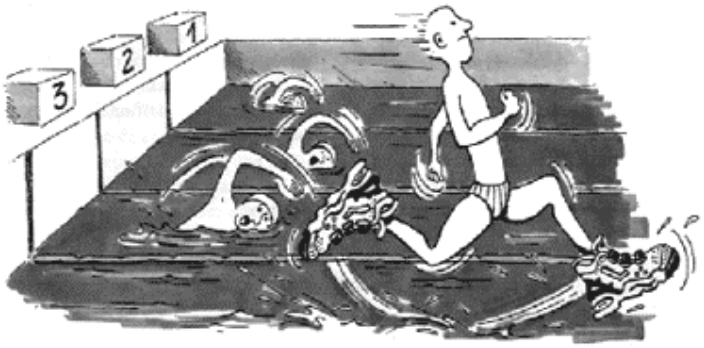
OR

Sydney: 7:00 AM – 8:00 AM; Berlin: 10:00 PM – 11:00 PM
Sydney 17:00, Berlin 8:00.

NOTE: If an interval is given, the entire interval must satisfy the constraints. Also, if morning (AM) or evening (PM) is not specified, but the times could otherwise be regarded as correct, the response should be given the benefit of the doubt, and coded as correct.

Appendix 5: Sample PISA Reading Literacy Items

Runners



For 14 years the Sports Medicine Center of Lyon (France) has been studying the injuries of young sports players and sports professionals. The study has established that the best course is prevention...and good shoes.

Knocks, falls, wear and tear...

Eighteen per cent of sports players aged 8 to 12 already have heel injuries. The cartilage of a soccer player's ankle does not respond well to shocks, and 25% of professionals have discovered for themselves that it is an especially weak point. The cartilage of the delicate knee joint can also be irreparably damaged and if care is not taken right from childhood (10-12 years of age), this can cause premature osteoarthritis. The hip does not escape damage either and, particularly when tired, players run the risk of fractures as a result of falls or collisions.

According to the study, soccer players who have been playing for more than ten years have bony outgrowths either on the tibia or on the heel. This is what is known as "soccer player's foot", a deformity caused by shoes with soles and ankle parts that are too flexible.

Protect, support, stabilize, absorb

If a shoe is too rigid, it restricts movement. If it is too flexible, it increases the risk of injuries and sprains. A good sports shoe should meet four criteria:

Firstly, it must provide exterior protection: resisting knocks from the ball or another player, coping with unevenness in the ground, and keeping the foot warm and dry even when it is freezing cold and raining.

It must support the foot, and in particular the ankle joint, to avoid sprains, swelling and other problems, which may even affect the knee.

It must also provide players with good stability so that they do not slip on a wet ground or skid on a surface that is too dry.

Finally, it must absorb shocks, especially those suffered by volleyball and basketball players who are constantly jumping.

Dry feet

To avoid minor but painful conditions such as blisters or even splits or athlete's foot (fungal infections), the shoe must allow evaporation of perspiration and must prevent outside dampness from getting in. The ideal

material for this is leather, which can be water-proofed to prevent the shoe from getting soaked the first time it rains.

Runners - Aspect: Interpreting Texts

Question Sample Question

What does the author intend to show in this text?

A. That the quality of many sports shoes has greatly improved.

B. That it is best not to play soccer if you are under 12 years of age.

C. That young people are suffering more and more injuries due to poor physical condition.

D. That it is very important for young sports players to wear good shoes.

Answer Sample Answer

The correct answer is D.

Runners - Aspect: Retrieving Information

Question Sample Question

One part of the article says, "A good sports shoe should meet four criteria."

What are these criteria?

Answer Sample Answer

To receive full credit, students must refer to the four criteria in italics in the text.

Runners - Aspect: Reflecting on Texts

Question Sample Question

Look at this sentence from near the end of the article. It is presented here in two parts:

"To avoid minor but painful conditions such as blisters or even splits or athlete's foot (fungal infections),…"

"…the shoe must allow evaporation of perspiration and must prevent outside dampness from getting in."

What is the relationship between the first and second parts of the sentence?

The second part

A. contradicts the first part.

B. repeats the first part.

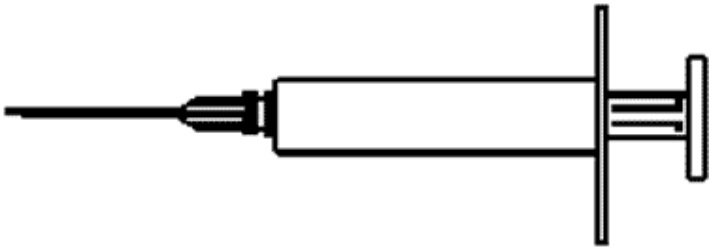
C. illustrates the problem described in the first part.

D. gives the solution to the problem described in the first part.

Answer Sample Answer

The correct answer is D.

Flu



ACOL Voluntary Flu Immunization Program

As you are no doubt aware the flu can strike rapidly and extensively during winter. It can leave its victims ill for weeks.

The best way to fight the virus is to have a fit and healthy body. Daily exercise and a diet including plenty of fruit and vegetables are highly recommended to assist the immune system to fight this invading virus.

ACOL has decided to offer staff the opportunity to be immunized against the flu as an additional way to prevent this insidious virus from spreading amongst us. ACOL has arranged for a nurse to administer the immunizations at ACOL, during a half-day session in work hours in the week of October 17. This program is free and available to all members of staff.

Participation is voluntary. Staff exercising the option will be asked to sign a consent form indicating that they do not have any allergies, and that they understand they may experience minor side effects.

Medical advice indicates that the immunization does not produce influenza. However, it may cause some side effects such as fatigue, mild fever and tenderness of the arm.

Who Should Be Imunized?

Anyone interested in being protected against the virus.

This immunization is especially recommended for people over the age of 65. But regardless of age, ANYONE who has a chronic debilitating disease, especially cardiac, pulmonary, bronchial or diabetic conditions.

In an office environment ALL staff are at risk of catching the flu.

Who Should Not Be Imunized?

Individuals hypersensitive to eggs, people suffering from an acute feverish illness and pregnant women.

Check with your doctor if you are taking any medication or have had a previous reaction to a flu injection.

If you would like to be immunized in the week of October 17 please advise the personnel officer, Anne Washington, by Friday October 7. The date and time will be set according to the availability of the nurse, the number of participants and the time convenient for most staff. If you would like to be immunized for this winter but cannot attend at the arranged time please let Anne know. An alternative session may be arranged if there are sufficient numbers.

For further information please contact Anne on ext. 5577.

Anne Washington, the personnel officer at a company called ACOL, prepared the information sheet on the previous two pages for ACOL staff. Refer to the information sheet to answer the questions which follow.

Flu - Aspect: Retrieving Information

Question Sample Question

Which one of the following describes a feature of the ACOL flu immunisation program?

A. Daily exercise classes will be run during the winter.

B. Immunizations will be given during working hours.

C. A small bonus will be offered to participants.

D. A doctor will give the shots.

Answer Sample Answer

The correct answer is B.

Flu - Aspect: Reflecting on Texts

Question Sample Question

We can talk about the content of a piece of writing (what it says).

We can talk about its style (the way it is presented).

Anne wanted the style of this information sheet to be friendly and encouraging.

Do you think she succeeded?

Explain your answer by referring in detail to the layout, style of writing, pictures or other graphics.

Answer Sample Answer

To receive full credit, responses must refer accurately to the text and relate style to purpose, consistent with "friendly and encouraging." The answer must

do at least one of the following: 1) refer to one of the features in detail and/or 2) use evaluative terms other than "friendly and encouraging." Opinion about whether Anne succeeded may be stated or implied.

Responses that refer accurately to the text and relate purpose to information and content (rather than style), consistent with "friendly and encouraging," receive partial credit. Opinion about whether Anne succeeded may be stated or implied.

Flu - Aspect: Interpreting Texts

Question Sample Question

This information sheet suggests that if you want to protect yourself against the flu virus, a flu shot is

A. more effective than exercise and a healthy diet, but more risky.

B. a good idea, but not a substitute for exercise and a healthy diet.

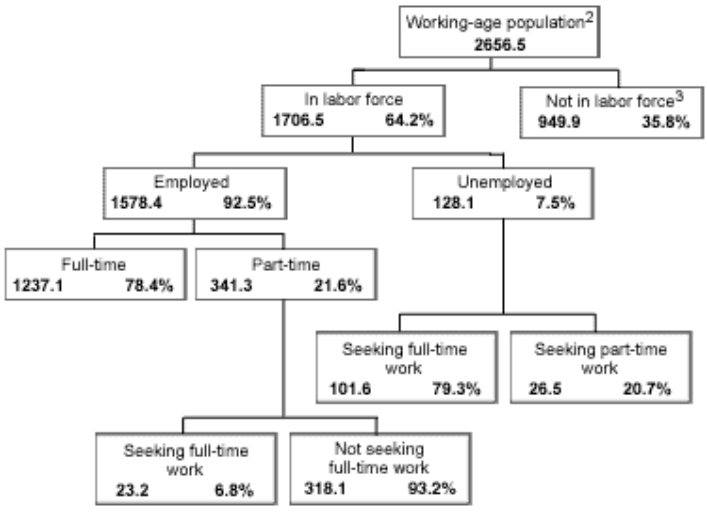
C. as effective as exercise and a healthy diet, and less troublesome.

D. not worth considering if you have plenty of exercise and a healthy diet.

Answer Sample Answer

The correct answer is B.

Labor



The tree diagram above shows the structure of a country's labor force or "working-age population". The total population of the country in 1995 was about 3.4 million.

Notes

1. Numbers of people are given in thousands (000s).

2. The working-age population is defined as people between the ages of 15 and 65.

3. People "Not in labor force" are those not actively seeking work and/or not available for work.

Use the information about a country's labor force on the opposite page to answer the questions below.

Labor - Aspect: Retrieving Information

Question Sample Question

How many people of working age were not in the labor force? (Write the number of people, not the percentage.)

Answer Sample Answer

To receive full credit, responses should indicate that the number in the tree diagram and the "000s" in the title/footnote have been integrated to produce the number 949,000. Approximations between 949,000 and 950,000, in numbers or words, were accepted. In addition, 900,000 or one million (in words or numbers) were accepted if accompanied by a qualifier such as "almost" or "about."

Correctly indicating the number in the tree diagram, but failing to integrate the "000s" in the title/footnote can achieve partial credit. In this situation, students would answer 949.9 in words or numbers.

Labor - Aspect: Interpreting Texts

Question Sample Question

In which part of the tree diagram, if any, would each of the people listed in the table below be included?

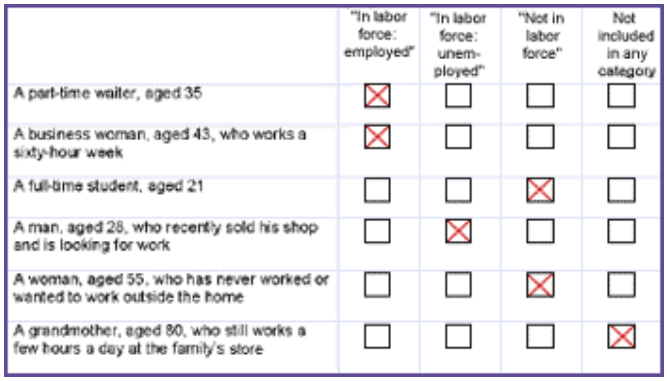
Show your answer by placing a cross in the correct box in the table.

The first one has been done for you.

	"In labor force: employed"	"In labor force: unemployed"	"Not in labor force"	Not included in any category
A part-time waiter, aged 35	☒	☐	☐	☐
A business woman, aged 43, who works a sixty-hour week	☐	☐	☐	☐
A full-time student, aged 21	☐	☐	☐	☐
A man, aged 28, who recently sold his shop and is looking for work	☐	☐	☐	☐
A woman, aged 55, who has never worked or wanted to work outside the home	☐	☐	☐	☐
A grandmother, aged 80, who still works a few hours a day at the family's store	☐	☐	☐	☐

Answer Sample Answer

The following is a correctly answered diagram.



	"In labor force: employed"	"In labor force: unemployed"	"Not in labor force"	Not included in any category
A part-time waiter, aged 35	☒	☐	☐	☐
A business woman, aged 43, who works a sixty-hour week	☒	☐	☐	☐
A full-time student, aged 21	☐	☐	☒	☐
A man, aged 28, who recently sold his shop and is looking for work	☐	☒	☐	☐
A woman, aged 55, who has never worked or wanted to work outside the home	☐	☐	☒	☐
A grandmother, aged 80, who still works a few hours a day at the family's store	☐	☐	☐	☒

To receive full credit, students needed to correctly identify each person's labor classification. Students who correctly identified 3 or 4 of the labor classifications received partial credit.

Copies of this publication can be obtained from:

DfES Publications
P.O. Box 5050
Sherwood Park
Annesley
Nottingham
NG15 0DJ

Tel: 0845 60 222 60
Fax: 0845 60 333 60
Minicom: 0845 60 555 60
Online: www.dfespublications.gov.uk

Produced by the Department for Education and Skills

ISBN 1 84478 765 6
Ref No: RR772
www.dfes.go.uk/research