RESEARCH AND EVALUATION DIVISION

www.ucles-red.cam.ac.uk

# Can a picture ruin a thousand words? Physical aspects of the way exam questions are laid out and the impact of changing them.

**Victoria Crisp and Ezekiel Sweiry**
**University of Cambridge Local Examinations Syndicate**

**A paper to be presented at the British Educational Research Association Annual Conference, Edinburgh, September 2003.**



**Contact details**
Victoria Crisp and Ezekiel Sweiry
Research and Evaluation Division,
University of Cambridge Local Examinations Syndicate,
1 Hills Road, Cambridge,
CB1 2EU.
☎ 01223 553805/553846
*FAX:* 01223 552700
crisp.v@ucles.org.uk or sweiry.e@ucles.org.uk

**Can a picture ruin a thousand words? Physical aspects of the way exam questions are laid out and the impact of changing them.**

**Abstract:**

Previous research suggests that physical aspects of the way an exam question is presented (e.g. the layout of the page, the diagrams or pictures used) can influence the way that students understand it and what kind of answer they think is required. In practice this means that sometimes certain features can lead students to answer in ways not intended by the question writer.

When reading a question, students form a mental representation of the task they are being asked to carry out. Certain aspects of a question such as diagrams or images are particularly salient and hence can come to dominate the mental representation that is formed. Therefore, subtle changes to these salient physical features of a question may affect how the question is understood.

This study set out to investigate the extent to which the nature of the physical features used in exam questions influence how students understand them. Questions based on past examination questions were retrialled in schools, along with modified versions of them. Changes in the students' performances between different versions of the same questions were analysed.

This constitutes a further stage in the collection of empirical evidence on the effects of features of exam questions on difficulty and validity. The information obtained from such research is used to inform training for question writers.

We will illustrate the findings using example questions and consider the implications for question writing.

**Introduction:**

Physical aspects of the way an exam question is presented can influence students' understanding of the task. For example, the location of question elements on a page can affect which information is perceived as more important and the quantity students think they need to write will be affected by the amount of answer space provided. Visual resources contained in exam questions, such as graphs, tables, diagrams, photographs and sketches, have sometimes been seen to influence students' understanding (Fisher-Hoch, Hughes and Bramley 1997). This paper will focus mainly on the use of these latter features.

Visual resources are sometimes included in order to test students' ability to use or interpret them, but they are more commonplace than this alone would warrant. However, there seems to have been little research into the effects of including diagrams in examination questions. Much of what is known about this comes from research on the influence of illustrations in instructional texts. Most of the research in this area has suggested that pictures have a positive influence on learning and retention, with text being remembered better when it is illustrated (Schnotz, 2002, Weidenmann, 1989, Ollerenshaw, Aidman, and Kidd, 1997). However, the main purpose of exam questions is to assess learning rather than teach and hence, in itself, this does not justify their use in exams.

Various other positive benefits reported by research on instructional texts may explain their use in examination questions. Graphics are thought to "simplify the complex" and "make the abstract more concrete" (Winn, 1989, p. 127). Peeck, (1993) makes a similar point when she writes that images "might help to clarify and interpret text content that is hard to comprehend" (p. 227). It is also argued that graphics can provide more information than can be explained in words (e.g. Stewart, Van Kirk and Rowell 1979). This could mean that in exams, including a clear illustration rather than a textual description could reduce the necessary length of questions.

In addition, images are generally believed to have a motivational role in the context of instructional texts (Peeck, 1993) which could apply equally to exam questions. Since examinations are stressful situations for most students, elements that trigger their interest or make a question look less daunting may be viewed as having an advantageous role.

On the other hand, in a review of studies on instructional texts, Levie and Lentz (1982) found that about 15% of them had observed no significant effects of including images. Peeck (1987) found that participants who read a text without a diagram were actually more motivated and more interested in reading more than those who read the same text accompanied by a poor diagram, suggesting that pictures are not always beneficial.

These failures of pictures to aid instruction have been explained in various ways; often as either a result of students' learning styles (as Ollerenshaw, Aidman, and Kidd, 1997 report) or due to students not processing illustrations adequately (Weidenmann, 1989). It is also pointed out that the apparent ease of processing an image may give a student the false impression that the image has been fully understood (Weidenmann, 1989). In addition, Winn (1989) warns text designers of making assumptions that all students will process a particular diagram in a particular way. This idiosyncrasy of interpretation is also implied by Elkins (1998), an art historian, who asserts that visual images do not provide meaning via an orderly set of signs in the same way that text does.

Perhaps the main possible negative effect of including pictures in exam questions is the risk that a picture may lead to the formation of a mental representation of the question that is not the one intended by the examiner. When a student reads a question, a mental representation is built up as a response to the text being processed. This representation is composed of images, concepts, emotions and the relationships between concepts, but not of actual words. It is also based on ideas that are already known to the reader (Johnson-Laird, 1981). The mental model will be the reader's own personal understanding of the text. Therefore students' mental representations of the text may not all be the same, perhaps emphasising certain aspects that seem particularly salient to them. Most of this process is unconscious and automatic, and involves the activation of related concepts in the mind.

Visual resources are thought to play a large role in the development of the student's mental model of the question and more emphasis will be placed on the ideas communicated by them than the ideas conveyed by the associated text. As Peeck (1987) states, "too much attention may be deployed to the illustrations themselves rather than to the accompanying text" (p.118). She also describes a previous study (Peeck, 1974) in which students were presented with a story that sometimes contained a mismatch of information between text and image. The students tended to choose the responses consistent with the pictures more frequently than the responses that would be indicated by the text, suggesting a dominating influence of the images.

There are a number of possible reasons for the apparent superiority of images over text. Firstly, it is thought that processing visual material requires less cognitive effort. According to Biedermann (1981) the general meaning of an image can usually be grasped in as little as 300 milliseconds. This may be because the elements of a visual source can usually be processed simultaneously, whereas text must be processed sequentially (Winn, 1987). This suggests advantages of using images to portray information in a rushed examination situation, since the overall meaning of a visual resource can be grasped more quickly than that of pure text.

Visual and textual materials may be processed in different cognitive systems or subsystems. Paivio's (1975) theory of dual-coding explains that the superiority of memory for images is a result of pictures being coded both as images and as their verbal labels whilst words are only encoded verbally. Thus the two representations of one item results in bias towards information gained from visual resources (Schnotz, 1993). Mayer (1989) argues further that the double coding of images facilitates the formation of the mental model since referential connections between the two representations will already be produced. However, there has also been opposition to this view and some have even claimed that images provided with text might be harmful since attention is split between the two forms of information which have to be integrated (Sweller, 1990).

In addition to the idea that placing information higher on a page generally makes it seem more valuable (Winn, 1987), there is also some evidence that visual resources are more likely to be read and processed *before* accompanying text. Kennedy (1974) discusses how "sometimes we read a label or caption before looking at the picture, but more often, probably, we notice the picture first and recognise the pictured object without any help from the accompanying words" (p. 7). It has been well documented that the first elements contained within a mental model will dominate and strongly influence subsequent elements (Gernsbacher, 1990). This is because the mental representation is started on the basis of the first element processed, and each subsequent piece of information is incorporated into the developing representation whenever possible. Hence the fact that images are likely to be processed first means they will be likely to dominate the representation.

If true, the argument that visual resources have a disproportionately large influence on the development of mental models has strong implications in examinations where students' ability to process material efficiently is already compromised by test anxiety (Sarason, 1988). This underlines the importance of ensuring that diagrams are accurate and unambiguous. In addition, irrelevant information included within a visual resource may result in the wrong information being used (although, of course, sometimes examiners may wish to test selection skills). Question writers also need to be aware that the salience of visual elements will affect students' understanding, and therefore that the key elements need to be the most salient ones.

These implications are demonstrated below using an example question.

T lymphocytes have protein receptors in their cell surface membranes. These T cell receptors are very similar in structure to antibody molecules. Each type of T cell receptor binds specifically to one type of antigen. Fig. 3.1 shows part of a cell surface membrane of a T cell with an antigen bound to a T cell receptor.

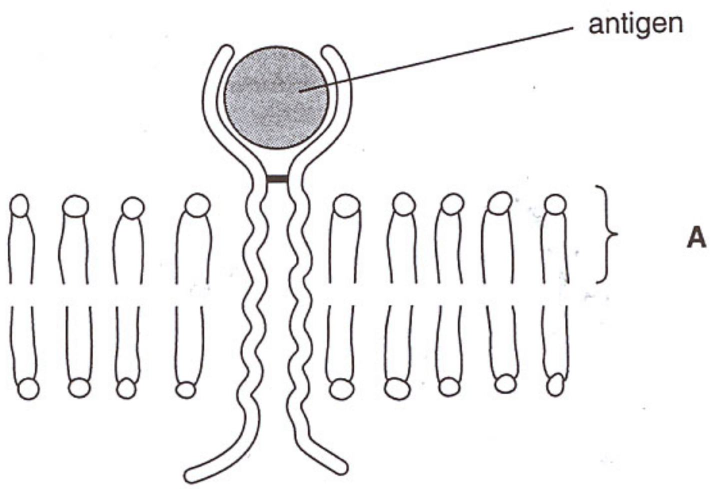



**Fig. 3.1**

(a) With reference to Fig. 3.1,

(i) name the molecule labelled **A** [1]

The question above is taken from a Biology AS-Level paper. The required answer to the question is 'phospholipid' or 'phospholipid molecule'. Given the probable dominance of diagrams within mental representations, and the fact that the diagram appears before the actual question, it is likely that students will study the diagram before reading the text. Students are asked to name the *molecule* that is labelled 'A'. However, the curly bracket used to label 'A' is often used to denote a layer or group of items, rather than a single item. It was found that a number of students wrote 'phospholipid layer', hence not scoring the mark. We would hypothesise that a disproportionate amount of attention is likely to have been paid to the diagram, and students would have developed a fairly strong idea of what the diagram showed and what was going to be asked in the question. This resulted in some students not paying sufficient attention to the crucial word 'molecule' in the question.

**Participants:**

525 students (269 boys and 256 girls) aged 16 years completed either of two versions of a science test paper. These students were all studying science at one of four secondary schools. 266 students completed version 1 of the test whilst the remaining 259 completed version 2.

The predicted GCSE grades of the students ranged from A* to G, with the majority being predicted around a C or a D. The table below shows the predicted grades of all the students involved. This distribution is fairly typical of the national school population.

| Grade | A* | A | B | C | D | E | F | G | unknown |
|---|---|---|---|---|---|---|---|---|---|
| Number of students | 3 | 35 | 88 | 130 | 145 | 83 | 23 | 8 | 10 |
| % of students | 0.6 | 6.7 | 16.8 | 24.8 | 27.6 | 15.8 | 4.4 | 1.5 | 1.9 |

**Test paper construction:**

The test included twelve questions. Six of these were included for the purposes of this study. The others were either control questions or included in this test in order to study other issues.

The questions for this study included graphical or layout elements that we predicted might have an influence on students' processing. One question was common to both versions of the test paper. For each of the other questions, two versions were constructed in order to investigate the effects of changes to visual resources on students' processing and responses.

The questions were compiled to form the two versions of a test paper. The versions were not counterbalanced since this would have required an impractical number of versions. Instead we aimed to make the groups of students completing each version as equivalent as possible in terms of gender, ability and school distribution by assigning the two versions of the test randomly among students.

**Procedure:**

The tests were carried out in exam conditions, in the students' normal classroom or laboratory during lesson time. In order to ensure validity, students were not told that we were researching the use of visual resources. Students were given forty minutes to complete the full test and were required to answer all questions.

Twenty-seven pairs of students were interviewed immediately after the test in a quiet room away from the classroom. The aim of the interviews was to gain an insight into students' use of visual elements when answering the test questions. Interviewees were given access to their papers during the interview to help prompt their recall. The students were interviewed in pairs in order to gain reactions from more students, to make the interviews less stressful, and because in previous work this has been shown to elicit more comments (e.g. Ahmed and Pollitt 2001). The interviews were semi-structured in nature.

**Ability of students:**

The predicted grades obtained for each of the students were converted into a score as shown below.

| Grade | A* | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|---|
| Score | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

These measures were used to calculate mean ability of students attempting each version of the test.

| Version | Mean | N | Standard deviation |
|---|---|---|---|
| 1 | 4.50 | 261 | 1.369 |
| 2 | 4.55 | 254 | 1.353 |
| Total | 4.52 | 515 | 1.360 |

As the table shows, the mean ability of the two groups was found to be very similar.

**Students total marks on the test:**

Our test seemed to discriminate pupils fairly:

| Grade | A* | A | B | C | D | E | F | G | unknown | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean mark | 36.33 | 35.06 | 30.57 | 27.35 | 24.17 | 21.95 | 20.13 | 17.38 | 27.10 | 26.25 |

**Analysis of data:**

As well as conventional marking, students' answers to some question parts were also coded by the kinds of responses given. Coding was carried out for question parts that were deemed to be likely to be influenced by the visual elements in the question.

Cross-tabulation analyses were run to see the effects of version on scores and on the kinds of answers given for the coded question parts.

Interviews were analysed for common comments about questions and some of these will be reported when discussing the questions.

**Results:**

**Question 1 – Insects**

In version 1 of this question, the phrase 'All numbers are in thousands' was positioned above the table as shown below. This format is common in tables published in books and magazines. Nevertheless, we expected that, in exam conditions, this might lead students to overlook this information and write the shortened value (i.e. 170) rather than the full value (i.e. 170,000). It was hypothesised that this might occur partly due to the layout of the various elements on the page, and that even if students started to read the text first their attention would then directed to the chart by the first sentence of the text and hence the second sentence may not be sufficiently processed. In including the same information within the table in Version 2 of the question, it was thought that it would be less readily overlooked due to its closer proximity to the numbers to which it refers and because it was emboldened.

Version 1

1 The chart below shows the average number of insects present per 10 km² in a woodland area over a number of months.

All numbers are in thousands.

| Name of Insect Species present in Woodland Area | May | June | July | August |
|---|---|---|---|---|
| Greenfly | 110 | 120 | 170 | 200 |
| Wasp | 15 | 16 | 15 | 16 |
| Ladybird | 6 | 5 | 3 | 2 |
| Dung Beetle | 10 | 11 | 11 | 11 |

Use the information above to answer the following questions.

**(a)** What was the average number of greenfly present per 10 km² in July?

............................[1]

Version 2

1 The chart below shows the average number of insects present per 10 km² in a woodland area over a number of months.

| | **Number of insects, in thousands** | | | |
|---|---|---|---|---|
| **Name of Insect Species present in Woodland Area** | **May** | **June** | **July** | **August** |
| Greenfly | 110 | 120 | 170 | 200 |
| Wasp | 15 | 16 | 15 | 16 |
| Ladybird | 6 | 5 | 3 | 2 |
| Dung Beetle | 10 | 11 | 11 | 11 |

Use the information above to answer the following questions.

**(a)** What was the average number of greenfly present per 10 km² in July?

............................[1]

Only answers of '170,000' and not '170' were given credit. In version 1, only 31.2% of students wrote the accepted answer of '170,000'. 65.0% of students in this version (almost all of those who didn't gain the mark) wrote '170'. This pattern was not significantly different in version 2 of the question where 33.6% of students gained the mark whilst 63.7% answered '170'.

From interviewing the students, it became apparent that most of those who answered incorrectly had not noticed the information that the numbers were in thousands. Indeed, most were startled to discover that their answer was incorrect, several accusing the question of being *'sneaky'*. One student said *'It's just trying to trick you, I don't think that's really what a science exam should be about, it should be testing your scientific knowledge not your ability to read a question'*. Although both presentation formats would probably not be problematic in normal settings, it seems that in a test, the students felt tricked. This student's comment raised an interesting question: is it fair under the stressful conditions of an examination to expect students to attend to particularly subtle aspects of questions?

Some students read only the first part of the text thoroughly and skimmed over the crucial piece of information regardless of its location, perhaps because it did not contain any key words. One student said *'I just read that bit [first sentence] and then looked at the chart and then looked at the question.'*

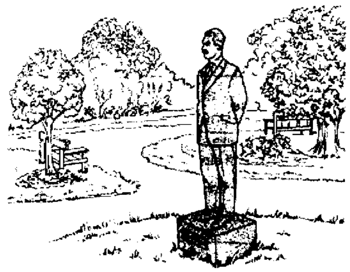
There were also some students who had noticed and taken in the information, but still did not write the required answer. This was because they did not think it necessary to write in the thousands when the numbers in the table weren't written out in thousands. One student said *'I just wrote down the answer that was in the box because I didn't think it was up to me to write it down because it's already said that it was in thousands.'* Another said *'I noticed the numbers in thousands but … I just kind of thought that well if they've only put 170 then that's what I would put'*.

The text containing the vital information appears to have been low in salience. Even when students did read the information, some did not expect that this information was required to answer the question given how the other values in the table are presented.

### Question 2 - Statue

This question was included partly to provide a question common to both versions of the test, and partly to investigate what use students make of a decorational diagram that isn't essential to a question. We wondered how much attention students would pay to the picture and how much they would think they needed to use it. Answers such as 'discoloured', 'material worn away' and 'bits broken off' were credited. Responses referring to a named feature such as nose or eyes being altered were also accepted. Hence, students did not necessarily need to use the diagram but it might help them gain the latter marking point.

2 The drawing shows a new statue made of sandstone. After some years the statue will look different because of weathering.



(a) Describe **two** ways in which the statue will look different because of weathering.

1.........................................................................................

.........................................................................................

2.........................................................................................

......................................................................................... [2]

Students tended to score well on this question. 85.5% of students (across both versions) scored both marks, whilst 12.4% scored one mark and only 2.1% scored none.

The picture could potentially be fulfilling a motivational role and also making the context of the question more concrete as well as providing features of the statue that can be named in describing how it would change.

There was a mixture of responses during interviews with regard to this sketch. Some said they found it useful, with slightly fewer stating that they did not find it of use or that it was unnecessary.

Positive comments included some impression of reassurance at having an illustration, and mention of how they could see that the statue had detailed features that might be lost. For example, one said, *'it gave you an idea of what to pick out…Like you would lose a lot of detail on the person'*.

Some students held quite strong views that the picture was unnecessary. For example, one claimed that *'You don't even need the picture. I mean if you say a statue it could be of anything, it works the same way… it shouldn't matter what the statue looks like.'* Another said that *'I just didn't look at it. I didn't bother with it because I looked at the question and it didn't ask anything about the picture… it doesn't really matter what the picture is.'*

One student said, *'Yes I think if I didn't have the diagram there I would probably use more science instead of just saying its features won't be so defined, I would probably have said acid rain would sort of wash it away'*.

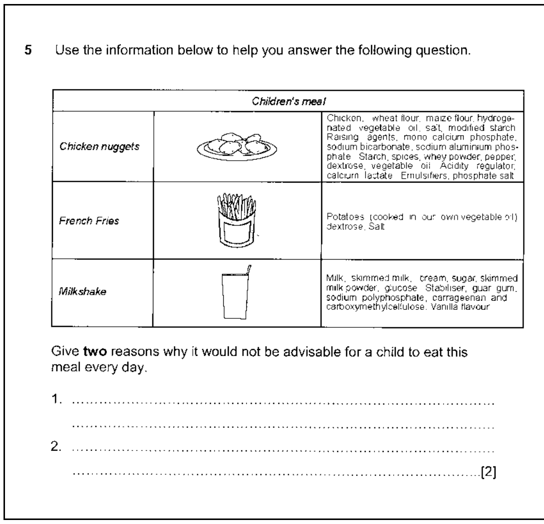
Interestingly, another commented that, *'The thing that it might do is that it might lead you in to look at the picture instead of the text so the answer might be wrong because you haven't read the text properly because the picture's avoiding you reading the text maybe. … It didn't help me answer the question.'*

The students' responses with regard to this question emphasise the point made earlier about how not all observers will use an image in the same way.

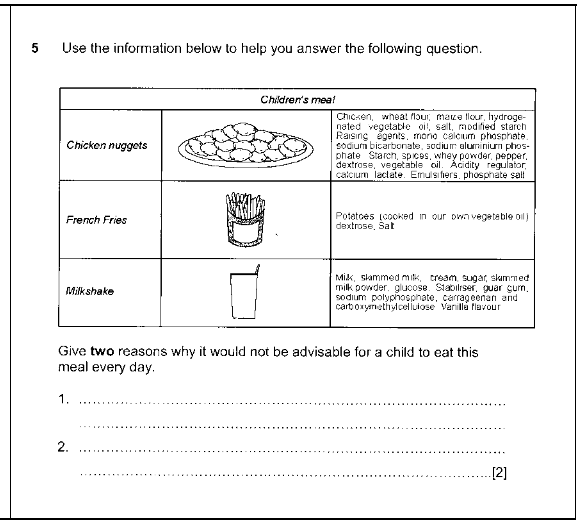
**Question 5 – Children's meal**

Version 1 acted as a control. We hypothesised that the salience of the large size of the portion of chicken nuggets in version 2 might dominate students' thinking and lead to related answers about overeating.

Version 1




Version 2




In the trial, no students wrote about the size of the portion or about overeating on either version of the question. The percentage of students scoring nought, one and two marks were very similar on the two different versions. For example, the percentage of students scoring two marks was 72.8% of those in version 1 and 72.3% of those in version 2.

There were two possible reasons why the number of chicken nuggets did not influence responses. Firstly, the image did not seem to be as salient to the students as we predicted it might be. Most students who answered version 2 of the paper who were interviewed said that they had not noticed the large portion size in the diagram. In addition, a lot of the comments during interview tended to suggest that students were not viewing these images as the actual meal that the child would eat, but just as a generic illustration of a meal of chicken nuggets, chips and milkshake, to indicate what they look like. For example, after the portion size had been mentioned by the interviewer, one student said, *'It doesn't really matter though, it just shows, illustrates what it is but it doesn't really matter how much there is'*. Several students made comments along the lines of *'I didn't actually take any notice of the picture, I only read the little caption next to it'* and *'I didn't really look at the pictures I went straight to the ingredients'*.

Secondly, some students who did notice the portion size did not use this in answering, perhaps not expecting an answer relating to amount of food to be worthy of credit in a science test. One student commented that *'You're thinking about it as in science so it* *would be like the content in it not the amount. The amount is nothing to do with science, is it?'*

In reference to this question another student made an interesting comment in favour of including images: *'the use of pictures isn't particularly useful in trying to answer the question, but it's quite daunting on the day if all you've got is text and you've just got to read it, so maybe a picture would calm your nerves'.*

In terms of salience of the image, more attention seems to be paid to the labels and ingredients information than to the picture, perhaps because the pictures are viewed as generic. Expectations that the quantity of food presented was relevant seemed to be low. It would be interesting to know whether students would have made more use of the picture if it had been less generic in nature (e.g. if it showed a child eating the meal).

**Question 6 - Products**

Some words can trigger more than one meaning in a student's mind. The word 'products' has a very specific meaning in chemistry but a more familiar meaning for students might be that of 'household products' and this might be triggered in an exam situation. Two versions of a question about chemical products were constructed based on different contexts. Version 1 was written in such a way that the likelihood that the wrong idea of 'product' would be triggered might be greater. This was done by using a context about toothpaste, and including a photograph that was intended to lead students' thought processes even more towards the everyday meaning of the word 'products'. Would the context, and particularly the photograph, dominate the students' mental model to such an extent that they were likely to activate the 'household' meaning of the word 'products', rather than the chemical one? In version 2 the context was more neutral.

Version 1

**6** After eating a meal, your mouth becomes very acidic. This acid can damage your teeth.

***[Part (a) omitted]***

Brushing your teeth with toothpaste will neutralise the acid. This will protect your teeth from damage.


***[Part (b) omitted]***

Some brands of toothpaste contain sodium carbonate.

**(c)** Three products are made when sodium carbonate reacts with hydrochloric acid.
What are they?

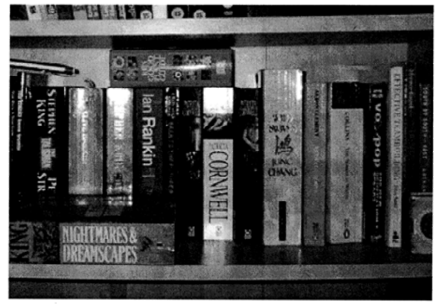
1..............................................................................................

2..............................................................................................

3..........................................................................................[3]

Version 2

**6** The paper in modern books contains slight traces of acid. The acid in the paper can make it slowly decay.




***[Part (a) omitted]***

One method of neutralising the acid in books is to use sodium carbonate.

***[Part (b) omitted]***

**(c)** Three products are made when sodium carbonate reacts with hydrochloric acid.
What are they?

1..............................................................................................

2..............................................................................................

3..........................................................................................[3]

In fact, 9% of students taking version 1 (toothpaste context) of this question gave answers such as 'shampoo' and 'soap' (words that it is assumed were activated by the context and photograph), while only 1.5% of students did this in version 2. The inclusion of the photograph in the toothpaste version of the question meant that the alternative understanding of the word 'products' tended to dominate some students' mental model of the question.

It appears that the toothpaste context appeared sufficiently scientific to many students, whereas this was not the case with the book version. For example, one student pointed out that sodium carbonate and hydrochloric acid could be used in shampoos. Another interviewee said that *'one [version 1] was more, makes sense, more scientific. That one [version 2] just seems off the point'*.

Even when both meanings of the word 'product' were active in students' consciousness, some interviewees still thought the household meaning of the word was the intended one. One student said that she *'thought it meant what was in the picture more than a chemical product'*.

## Question 9 – Components of a balanced diet

This question simply requires students to name the two other components of a balanced diet, the given one being 'proteins'. Version 1 used corned beef as an example and included a photograph. This was done in order to see whether this would make students more likely to answer with example *foods* rather than *components* e.g. potatoes rather than carbohydrates. It was thought that the inclusion of vegetables in the label on the tin illustrated might also trigger ideas about other foods. In version 2 no example or image was included.

**Version 1**

9 A meat canning company advertises its corned beef as high protein and easy to digest.


(a) Corned beef is rich in protein, which is one of the major components of a balanced diet.

Name the other **major** components of a balanced diet.

1 ..............................................................

2 .............................................................. [2]

**Version 2**

9 (a) Protein is one of the major components of a balanced diet.

Name the other **major** components of a balanced diet.

1 ..............................................................

2 .............................................................. [2]

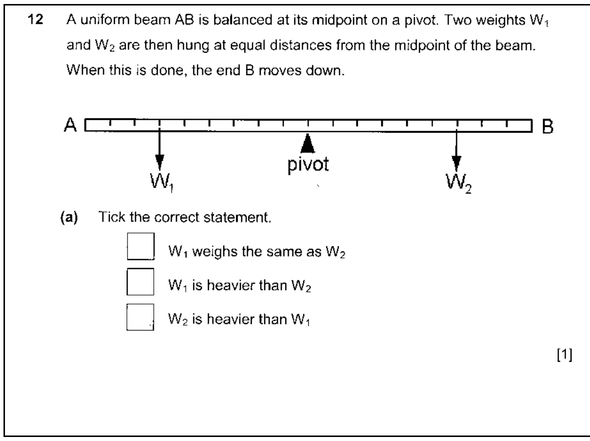
There was actually very little difference between responses in the two versions, both with respect to actual marks, and the extent to which students wrote names of foods rather than components. Students taking version 1 of the question were slightly more likely to score 1 mark (52.9% versus 50.8%) and 2 marks (35.6% versus 33.3%). There was no difference in the number of students who gave names of foods as their answers, suggesting the photograph of the corned beef had no effect.

While comments from the interviewees clearly show that the photograph was attended to, students were also aware that it was irrelevant. One student remarked that *'the picture doesn't do anything... it's not really useful for the question'*. Another said *'A lot of the time the question … they include an example would have a picture just confuses you more. It just distracts you, distracts you from actually answering the question'*.

## Question 12 - Balance

In version 1 of this question the text states that end B moves down when weights are attached to the beam, but the beam is horizontal in the diagram. In version 2, the beam was slanted in order to reflect the correct answer. This was done in order to investigate whether students might overlook the textual information since the diagram appears to be able to supply the answer and also whether students expect technical diagrams of this nature to be relevant and informative.

Version 1



12 A uniform beam AB is balanced at its midpoint on a pivot. Two weights $W_1$ and $W_2$ are then hung at equal distances from the midpoint of the beam. When this is done, the end B moves down.

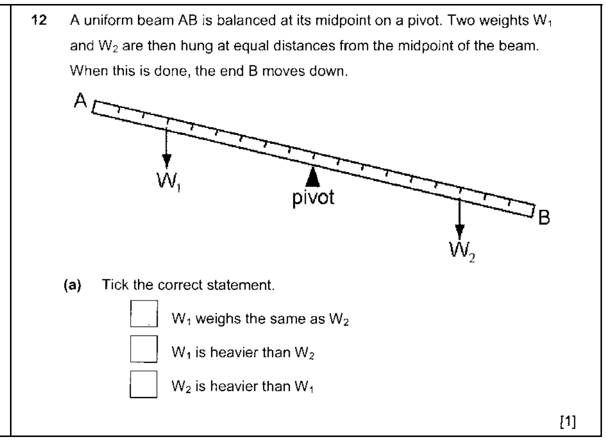


(a) Tick the correct statement.

- $W_1$ weighs the same as $W_2$
- $W_1$ is heavier than $W_2$
- $W_2$ is heavier than $W_1$

[1]

Version 2



12 A uniform beam AB is balanced at its midpoint on a pivot. Two weights $W_1$ and $W_2$ are then hung at equal distances from the midpoint of the beam. When this is done, the end B moves down.



(a) Tick the correct statement.

- $W_1$ weighs the same as $W_2$
- $W_1$ is heavier than $W_2$
- $W_2$ is heavier than $W_1$

[1]

There was a considerable difference in the number of correct answers in the two versions (80.4% in version 1 and 98.8% in version 2). 16% of the students taking version 1 thought the two weights were equal, suggesting that more attention had been paid to the diagram than the last sentence of the text. This was the only question for which there was a statistically significant difference ($F = 319.09$, $p < 0.01$) between marks on the two versions of the question. This difference was also reflected in the interviews, where one student said *'you see I didn't read the question on that, I just looked at the diagram, the picture'*.

Students clearly expected the diagram to reflect the answer. One student said *'Oh that actually says it moves down but it hasn't got it in the picture. That confused me because it's got … it's got the text in one thing, saying one thing and the picture saying they're level'*. In a similar vein, another said *'I just didn't read the last bit or I might have read it and then ignored it and then I looked at the diagram and then went from the diagram because from earlier questions the diagram was important. You always assume that the diagram is what they're saying in the question, the question comes after it.'*

**Discussion:**

Predicting the effect of visual resources (and the elements within them) on the way examination questions are answered is a complex task. Nevertheless, an analysis of the example questions provided in this paper, along with others the authors have studied, suggest that two variables in particular play a decisive role in determining the effect visual resources may have on the way examination questions are understood.

The first of these is the relative salience of the key element(s) of the visual resource. In order for an element of a diagram to be of influence, it must at the very least be processed. Yet mere processing is not always sufficient, as, for example, a question dense in information may result in key elements not being maintained in memory. The more salient an aspect of a question is relative to the rest of the question, therefore, the larger the part it is likely to play in the students' mental model of the question. Salience in exam questions can be increased by increasing the size of the relevant element, by using bold, and by including the element at the beginning of the question.

However, relative salience on its own is not enough. The student must also believe that the element is relevant to the answer. There are a number of factors that may determine whether a student believes an element to be relevant. Most important is past examination experience, which will provide students with expectations (Sweiry *et al,* 2002) regarding under what circumstances diagrams are relevant, and therefore how much attention to pay to them. These experiences will also help students determine whether different types of answers triggered by the visual resources are relevant. The type of image used and the context within which the question is set may also determine the extent to which the significant elements within it are deemed relevant.

In addition, the two variables (salience and relevance) interact, as the greater the relative salience of an aspect of a visual resource, the more likely the student is to deem it relevant. In the case of an element of very high salience, students would have to have very low expectations of its relevance to exclude using it in answering a question.

The two variables were found to be very reliable at predicting how images would be used when applied to the questions contained within this paper. In the example question about T lymphocytes, the diagram is large and hence salient in comparison to the question text. The method used for labelling 'A' (a curly bracket rather than a line) is also salient in the context of an exam, and leads to expectations about the likely question before it is actually read. Students will likely see relevance in the fact that 'A' is labelled in a different way to the 'antigen'. This high salience and perceived relevance of the curly bracket resulted in some students answering 'phospholipid layer'.

In question 5, the key element is the number of chicken nuggets. Evidence from the interviews suggests that this feature was of average salience – while the picture was clear, students paid more attention to the headings and ingredients, and were likely to view the diagrams as generic. In addition, relevance was deemed to be low, as students did not expect that the number of chicken nuggets (and hence the idea of over-eating) would be relevant in a science paper.

The photograph of the toiletries in question 6 was highly salient (due to its size and unexpected nature), and was also deemed relevant due to its association with ideas about chemistry. This was not the case with the 'books' version of the question.

In question 1, the key phrase 'all numbers are in thousands' was low in relative salience regardless of where in the question it was displayed. This is because the numbers contained within the table were not shown in thousands. Therefore, it is possible that the phrase was not processed at all. Even if it was attended to, students may have deemed the relevance of the phrase to be low, as the question appeared to simply require a figure to be located directly from the table. Hence, in both versions, many students failed to write their answer in thousands.

Evans and Over (1996) distinguish between two kinds of reasoning. What we may call naturalistic reasoning is innate and is used in everyday functioning, whilst formal reasoning is logical and learnt. It is the latter reasoning that exams seek to assess. Evidence from students' comments about question 2 (statue question) suggest that scientific diagrams are more likely to encourage formal reasoning, and naturalistic pictures are more likely to elicit naturalistic reasoning. The technical diagram in question 12 (balance) was more likely to be taken seriously; students thought the diagram would accurately reflect the answer.

**Conclusion:**

The use of visual resources in examination papers can serve various positive purposes. However, the effects of images are somewhat unpredictable and hence caution is required. Considering the salience and perceived relevance of visual elements can aid the prediction of their effects.

Students have strong expectations regarding the purpose of visual resources, and will react in ways they have learned based on their experiences. Students also have expectations based on the type of image used in a question, which may influence the kind of reasoning that they use. In summary, because students will hold strong expectations about images and how to use them, examiners would be advised to think very carefully before including anything unexpected in an exam question as students will react in ways almost predetermined by their past experience.

**References:**

Ahmed, A. & Pollitt, A. (2001) *Improving the validity of contextualised questions: a paper presented at the British Educational Research Association Annual Conference*, Cambridge: University of Cambridge Local Examinations Syndicate, Available from http://www.ucles-red.cam.ac.uk/conferencepapers.htm

Biedermann, I. (1981) On the semantics of a glance at a scene, in M. Kubovy & J. R. Pomerantz (Eds), *Perceptual organization,* Hillsdale, NJ: Erlbaum.

Elkins, J. (1998) *On pictures and the words that fail them*, Cambridge: Cambridge University Press.

Evans, J. St. B. T. & Over, D. E. (1996) *Rationality and Reasoning*, Hove: Lawrence Erlbaum.

Fisher-Hoch, H., Hughes, S. & Bramley, T. (1997) *What makes GCSE examination questions difficult? Outcomes of manipulating difficulty of GCSE questions: a paper presented at the British Educational Research Association Annual Conference*, Cambridge: University of Cambridge Local Examinations Syndicate, Available from http://www.ucles-red.cam.ac.uk/conferencepapers.htm

Gernsbacher, M. A. (1990) Language comprehension as structure building, Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. (1981) Mental models of meaning, in A. K. Joshi, B. L. Webber & I. A. Sag (Eds), *Elements of Discourse Understanding*, Cambridge: Cambridge University Press.

Kennedy, J. M. (1974) *Psychology of picture perception*, San Francisco: Jossey-Bass.

Levie, H. W., & Lentz, R. (1982) Effects of text illustrations: A review of research, *Educational Communication and Technology Journal,* 30, pp. 195-232.

Mayer, R. E. (1989) Systematic thinking fostered by illustrations in scientific text, *Journal of Educational Psychology,* 81 (2), pp. 240-246.

Ollerenshaw, A., Aidman, E. & Kidd, G. (1997) Is an illustration always worth ten thousand words? Effects of prior knowledge, learning style and multimedia illustrations on text comprehension, *International Journal of Instructional Media*, 24 (3), pp. 227-238.

Paivio, A. (1975) Imagery and Long-term Memory, in A. Kennedy & A. Wilkes (Eds), *Studies in Long Term Memory*, London: Wiley.

Peeck, J. (1974) Retention of pictorial and verbal content of a text with illustrations, *Journal of Educational Psychology,* 66, pp. 880-888.

Peeck, J. (1987) The Role of Illustrations in Processing and Remembering Illustrated Text, in D. M. Willows & H. A. Houghton (Eds), *The Psychology Of Illustration, Volume 1 Basic Research*, New York: Springer-Verlag.

Peeck, J. (1993) Increasing picture effects in learning from illustrated text, *Learning and Instruction,* 3, pp. 227-238.

Sarason, I. G. (1988) Anxiety, self-preoccupation and attention, *Anxiety Research*, 1, pp. 3-8.

Schnotz, W. (1993) Some remarks on the commentary on the relation of dual coding and mental models in graphic comprehension, *Learning and Instruction,* 3, pp. 247-249.

Schnotz, W. (2002) Towards an integrated view of learning from text and visual displays, *Educational Psychology Review,* 14 (1), pp. 101-120.

Stewart, J. H., Van Kirk, J., & Rowell, R. (1979) Concept maps: A tool for use in biology teaching, *The American Biology Teacher,* 41, pp. 171-175.

Sweiry, E., Crisp, V., Ahmed, A. & Pollitt, A. (2002) *Tales of the Expected: The influence of students' expectations on exam validity: a paper presented at the British Educational Research Association Annual Conference*, Cambridge: University of Cambridge Local Examinations Syndicate, Available from http://www.ucles-red.cam.ac.uk/conferencepapers.htm

Sweller, J. (1990) Cognitive processes and instructional procedures, *Australian Journal of Education,* 34 (2), 125-130.

Weidenmann, B. (1989) When good pictures fail: An Information-Processing Approach to the effect of illustrations, in H. Mandl, & J. R. Levin, (Eds), *Knowledge acquisition from text and pictures, Advances in Psychology 58*, Amsterdam: Elsevier.

Winn, W. (1987) Charts, graphs, and diagrams in educational materials, in D. M. Willows & H. A. Houghton (Eds), *The Psychology of Illustration, Volume 1 Basic Research*, New York: Springer-Verlag.

Winn, W. (1989) The design and use of instructional graphics, in H. Mandl, & J. R. Levin (Eds), *Knowledge acquisition from text and pictures, Advances in Psychology 58,* Amsterdam: Elsevier.