# Investigating gender differences in the science performance of sixteen-year-old pupils

John F Bell

# University of Cambridge Local Examination Syndicate

A paper presented at BERA, University of Sussex, Brighton, 2 – 5 September 1999

# Disclaimer

The opinions expressed in this paper are those of the author and are not to be taken as the opinions of the University of Cambridge Local Examinations Syndicate.

## Note

This research is based on data collected by the University of Cambridge Local Examinations Syndicate for Oxford, Cambridge and RSA examinations (OCR).

# Contact details

John F Bell, RED, UCLES, 1 Hills Road, Cambridge, CB1 2EU Tel. 01223 553849, Fax. 01223 360104, E-mail: bell.j@ucles.org.uk

## Abstract

Although the overall grades for General Certificate of Secondary Education (GCSE) science would suggest there were no gender differences in performance, when the examination is analysed at the question level the traditional gender stereotypes of boys outperforming girls in physics and girls outperforming boys in biology are still present. The paper will consider the performance on individual question parts. The question parts were placed into one of three categories: *Retrieving scientific information, Interpreting presented information*, and *Applying science concepts*. The paper continues with consideration of the memory processes involved in answering questions from the category *Retrieving scientific information* and concludes with a discussion of the relationship between the processes and the gender differences in science performance.

## Introduction

In this paper, gender differences in science performance on question parts that occur in the written components of a science GCSE examination will be considered. To interpret the pattern of differences the Assessment of Performance Unit's framework (Archenhold, *et al.*, 1988) will be considered.

In the next section of this paper, a modified framework will be described. This is followed by some background information about the GCSE data used in the paper. Then there is a section that is a detailed analysis of a sample of data drawn from the entry of the GCSE examination. The memory processes used in answering the question parts from the *Retrieving Scientific Information* category will be considered. This paper concludes with a brief discussion of the issues raised.

## **Classification of questions**

To investigate why there are gender differences in performance, it would seem sensible to consider the processes involved in answering questions. The first step in making this interpretation is to classify the questions. In previous work in this area, the APU assessment framework was considered (Bell, 1997b). This resulted in a modified framework with three categories, which are given in Table 1. Note that the classification used in this paper has been developed for the purposes of research. Although these have been referred to as question categories, it is more correct to consider them as a classification of the approaches used to answer questions. The new category *Retrieving scientific information* includes questions in which candidates are asked to give the standard name for some scientific feature, or a standard meaning of some standard terms or an explanation of a common theory. The answers to these questions only require remembering an answer and only a minimal understanding of the information. In Bell (1997b) this was referred to as *Recalling scientific information*. This change has been made so the name includes two different memory processes: recognition and recall.

ID	Category	Description
Ι	Interpreting presented	Question in which all the information needed to answer it is
	information	presented or it does not require ideas met in science lessons
А	Applying Science Concepts	Explicitly requires an understanding of science concepts
R	Retrieving scientific	Requires the use of memory to retrieve the name of a
	information	process or feature or the meaning of a name

## Table 1: Question Categories

All questions require some minimal conceptualisation of the phenomena under discussion. Questions in the category *Applying Science Concepts* are based on the necessity to use (not the mere possibility of using) ideas that a pupil is likely only to have met in science lessons. The questions from the category *Applying Science* explicitly require an understanding of taught science concepts. They require the application of such ideas in situations representative both of normal school science and of more wide ranging settings. The category *Interpreting presented information* includes questions that test the procedural skills involved in science but do not include questions that additionally require the retrieval of science concepts (which are placed in the category *Applying Science Concepts*). Also included in category *Interpreting presented information* necessary to answer them is given in the question.

For some questions, it is difficult to determine which of the three categories should be applied. For the purposes of analysis the categorisation of the questions was carried out for parts of questions for which a mark was explicitly awarded. There are two features to note. For whole examination questions, all three categories could be used. In addition, inspecting the mark schemes reveals that within question parts marks could be awarded to different categories (for example, in *Applying Science Concepts* a mark may be awarded for retrieving the concept and further marks for the actual application). However, without remarking the scripts, it is not possible to separate these implicit parts. In addition there is a problem

deciding, for some question parts, whether it is possible that the candidates had memorised an answer or whether they had applied a concept.

The proposed relationship between difficulty and question categories is presented in Figure 1. It is assumed that there are two dimensions relating to the question: the amount of scientific procedural skill required to answer the question and the amount of knowledge required to answer the question. *Retrieving scientific information* requires only low levels of scientific procedural skills but varying amounts of scientific knowledge, and questions in this category would be placed along the bottom of the figure. *Interpreting presented information* requires low levels of knowledge but varying levels of procedural skills and should be placed along the left of the figure. *Applying Science Concepts* requires both skills and knowledge and should be placed in the central area of the figure. The design of the figure emphasises the fact that the questions cannot be ordered in difficulty by category. It also illustrates the expected performance of differing types of candidates: an intelligent but lazy candidate would tend to have high levels of skills but would have low levels of knowledge from lack of revision so would be able to answer questions on the left side of the figure in the *Interpreting presented information* category. A less intelligent but diligent candidate would tend to be able to answer questions along the bottom of the figure in the *Retrieving Scientific Information* category.



Figure 1: Relationship between difficulty and question categories

Figure 1 also emphasises the fact that the overall difficulty of questions may vary within the categories. An *Applying Science Concepts* question may be easier than a *Retrieving Scientific Information* question when the level of recall and skill is lower. The questions used in an examination could result in any order of difficulty for the three categories.

Although this model has three categories of questions, the remainder of this paper concentrates on only one category, *Retrieving scientific information*. Each question part for all of the nine papers of a GCSE examination was classified as being a 'retrieval' question or not. It is relatively easy to classify some question parts because the only possible way of answering them is by retrieving scientific information. There are other question parts for which the decision is less straightforward, requiring a decision about whether retrieval was expected or whether the candidates would be expected to generate an answer by applying more general scientific concepts. This is complicated by the fact the several marks are awarded for some question parts. Inspection of the mark schemes revealed that marks could be awarded for parts of the response that could be allocated to different categories; e.g. one mark could be gained by recalling some scientific information and the second mark for applying it. It should be recognised that the objective of this exercise was to investigate whether gender differences occur. This can be demonstrated by finding questions where the only possible category is *Retrieving scientific information*.

There are two reasons for concentrating on the category *Retrieving scientific information*. Firstly, the cognitive processes involved in answering such questions are relatively straightforward and, secondly, the retrieval of appropriate scientific information is a vital requirement in answering questions in the category

*Applying Scientific Concepts.* This importance was demonstrated by Zeitoun (1989) who carried out a study to investigate the relationship between the achievement on a test of some abstract concepts in molecular genetics, prior knowledge, formal reasoning, ability and gender. He found that prior knowledge and formal reasoning ability play a major role in students' achievement on a test of abstract concepts; and the effect of students' prior knowledge on this achievement seemed to exceed that of formal reasoning ability.

## Background Information for the 1997 Nuffield Co-ordinated Sciences GCSE Double award

The data used to investigate the gender difference in retrieval and non-retrieval comes from the summer 1997 administration the Nuffield Co-ordinated Sciences General Certificate of Secondary Education Double Award Examination offered by the Midland Examining Group (now OCR). This syllabus has subsequently been withdrawn from use. The GCSE is an examination that is taken by most sixteen-year-olds in England Wales and Northern Ireland. Most GCSE examinations are effectively single award. In science, there are single and double award examination compared with a single award (in practice, it is usually slightly less). They receive two identical grades in recognition of the additional material covered. Double award GCSE science examinations are taken by the majority of sixteen-year-old pupils in England, Wales and Northern Ireland. More details of the uptake of GCSE science examinations can be found in Bell (1998) and (Bell, in prep.).

In 1997, assessment was by means of course work for the Attainment Target Sc1: Scientific Investigations, and terminal written examinations for Sc2: Life and Living Processes, Sc3: Materials and their Properties, and Sc4: Physical Processes. (Loosely speaking the four attainment targets can be considered as practical science, biology, chemistry and physics.) There were three tiers of entry for written examinations. Tiering is used to allow questions to be targeted for different levels of scientific attainment. Each tier contained questions set on specific National Curriculum levels as indicated in Table 2. This meant that there were nine written examinations for this syllabus. The Attainment Target Sc1 was assessed by the teachers within schools and is not tiered. This component is not considered in detail in this paper. The most able candidates were entered for the Further Tier. The entry for this particular syllabus tended to attract candidates from schools with a high proportion of candidates able to tackle the further tier.

Tier of entry	Components	Grades Available
Basic	1 (Sc2), 2 (Sc3), 3 (Sc4)	G-D (C)
Central	4 (Sc2), 5 (Sc3), 6 (Sc4)	F-B (A)
Further	7 (Sc2), 8 (Sc3), 9 (Sc4)	D-A*

Table	2: Tie	r structure	for	1997	GCSE	Nuffield	Co	-ordinated	Sciences:	Double	Award	1772
Lable	<b></b>	bulucture	101	1///	<b>UCDL</b>	1 (unitional	$\mathbf{v}$	orumateu	Delences.	Double		

(The award of grade C for the Basic tier and Grade A for the central tier was considered in exceptional cases only)

For each tier, there are three components (examination papers). In addition, the last component assesses coursework and is not tiered. There is an overlap between questions and question parts between tiers. The central tier paper 5, which assesses Sc3, includes question parts from both paper 2 and paper 8. Usually when an overlap occurs the question in the higher tier paper will have additional parts compared with the question in the lower tier. Figure 2 is the cumulative grade distribution by sex for Nuffield Co-ordinated Sciences Double Award. For this particular science syllabus, male candidates obtain better results than female candidates but the difference is small.





Figure 2, however, does not tell the whole story. In Figure 3, the distributions are presented for each attainment target. There are several interesting features displayed in this Figure. Firstly, candidates tend to obtain higher grades for Sc1 compared with the other attainment targets. There is a large gender difference in performance in favour of female candidates for Sc1. For Sc2, the biology-related attainment target, there is a smaller gender difference in favour of females and for Sc3, the chemistry related target, there is no evidence of an important gender difference. Finally, there is a large gender difference in performance favouring male candidates for the attainment target Sc4, the physics related attainment target. These results are consistent with the usual gender difference in performance found in science. A fuller description of performance on the previous year's attainment targets can be found in Bell (1997a).



Figure 3: Grade distributions by sex for each attainment target for Nuffield Co-ordinated Science Double Award

Before considering the gender differences for the question parts for each tier, it is necessary to investigate the levels of performance of the sexes on the tiers. This was done by considering the mean GCSE scores of the candidates across all the GCSE subjects they took. There are differences in mean GCSE performance by sex. Female candidates perform better than male candidates on subjects such as English and modern languages and the uptake of subjects varied between the sexes (Bell, 1998). To allow for these differences the percentage rank of mean GCSE score by sex for each candidate was calculated. A male candidate with a percentage rank of 80 would have a mean GCSE score such that only 20% of the male candidates would have a better score. The means of these two measures by sex and tier are presented in Table 3.

			Mean GCSE		Percentage Rank	
Tier	Sex	Ν	Mean	Std Dev.	Mean	Std Dev.
В	Female	1869	3.3	0.9	25	15
В	Male	1879	2.8	0.9	25	15
С	Female	7116	5.1	0.9	60	19
С	Male	6800	4.6	1.0	59	19
F	Female	3873	6.7	0.7	88	10
F	Male	4095	6.2	0.8	87	11

Table	3:	Mean	GCSE	and	percentage	rank	bv	tier	and	sex
	•••		0001		percenter		~ .			

Although for each tier the mean GCSE measure suggests that the male candidates' performance tends to be weaker than the female candidates', the percentage rank suggests that there is no difference. The fact that the percentage rank is similar suggests that any differences in the performance on question parts are not the result of differential uptake of candidates in the tiers.

For the purposes of further analysis, the question parts in all three tiers of the examination were classified into those that required only the retrieval of scientific information and those that did not. The distribution of marks for these types of questions is presented in Table 4. The main feature of this table is that the proportion of retrieval questions is lowest in the further tier.

Table 4: Distribution of	marks for retrieval and	d non-retrieval ques	stion parts by pa	per, tier, and
attainment target				

(recentages of marks anocated to each type by paper is given in the brackets)									
	Attainment target	Retrieval	Non-retrieval	Total					
Paper 1	Sc2	42 (47)	48 (53)	90					
Paper 2	Sc3	28 (37)	47 (63)	75					
Paper 3	Sc4	32 (43)	43 (57)	75					
Basic Tier	Sc2, Sc3, Sc4	102 (43)	138 (57)	240					
Paper 4	Sc2	43 (48)	47 (52)	90					
Paper 5	Sc3	38 (42)	52 (58)	90					
Paper 6	Sc4	42 (47)	48 (53)	90					
Central Tier	Sc2, Sc3, Sc4	123 (46)	147 (54)	270					
Paper 7	Sc2	34 (32)	71 (68)	105					
Paper 8	Sc3	29 (28)	76 (72)	105					
Paper 9	Sc4	41 (39)	64 (61)	105					
Further Tier	Sc2, Sc3, Sc4	104 (33)	211 (67)	315					

(Percentages of marks allocated to each type by paper is given in the brackets)

## An analysis of the Question Parts of the written components

Random samples of approximately 250 candidates were selected and the script for each written component was found. The data at the individual question part level were entered. In addition, for each question, a short descriptive synopsis was written. Usually only the total mark on a component is entered and used by the computer system in the administration of the examination. For each paper, two sets of analyses were carried out. Firstly, gender differences were investigated at the individual question part level and, secondly, the overall performance on retrieval and non-retrieval questions was considered. For each question part, a t-test was carried out by sex to identify those questions for which the gender difference was significant at the 5% level. Because of the large number of t-tests carried out some of the significant differences will have occurred by chance.

In Figures 4-6, the t-values generated by carrying out t-tests on the difference between the female mean score and the male mean score for each question part on each tier are presented. It should be noted that some question parts occurred in more than one tier. A positive t-value occurs when the female mean score is higher than the male mean score and vice versa.

In Figure 4, the results for the basic tier are presented. For most of the question parts that assess Sc2, the value of the t-value is positive indicating that the mean score is higher for females compared with males. The pattern for Sc3 is different. Most of the t-values are small and there is no evidence of a systematic gender difference. Finally, the t-values for most of the question parts assessing Sc4 are negative indicating a better performance by male candidates.

For paper 1, the question parts with significant sex differences are listed in Table 5. In addition, a classification of the question into retrieval and non-retrieval and a brief synopsis or description of the question have been included. Question part Q1A has been included although the difference is not significant because it illustrates an interesting pattern in performance.

Quest. ID	t-value	Туре	Synopsis
Q1C	3.78	R	Arrange three stages of birth into the correct order
Q9E	2.56	R	Given some information, explain why an electric heart is used on the
			left hand side of the heart
Q1D	2.55	R	Name the structure that joins the placenta to the baby
Q1B	2.54	R	Complete by adding words a paragraph about sexual reproduction
Q2C	2.02	Ν	Choose one organism from the four food chains which is a producer
Q4A	1.98	R	Complete a table listing sense organs and what they detect
Q1A	-1.76	R	Label four parts of the male reproductive system
Q10C	-2.97	R	Suggest two reason why few weeds or grasses live in the old woodland

Table 5: Question parts with significant gender differences for Paper 1 – Basic Tier (Sc2: Life and living processes)

Further details of the questions with statistically significant sex differences in paper 2 are presented in Table 6. There were only three significant differences: one in favour of female candidates and two in favour of male candidates.

# Table 5: Question parts with significant gender differences for Paper 2 – Basic tier (Sc3: Materials and their properties)

Quest. ID	t-value	Туре	Synopsis
Q4A	2.46	R	What is a polymer?
Q8C	-2.29	Ν	Given information, select which liquid in a table is best for neutralising a bee sting
Q5A	-2.51	R	Complete a sentence about fuel burning



# (a) Paper 1 (Sc2: Life and living processes)



(b) (Sc3: Materials and their properties







Table 7 lists the question with statistically significant t-values for the physics-based attainment target Sc3. The questions with the largest negative t-values tended to relate to astronomy, electricity and mechanics.

Table 7: Question parts with significant gender differences for Paper 3 – Basic Tier (Sc4: Physical processes)

Quest. ID	t-value	Туре	Synopsis
Q2B	3.28	R	What two colours for light bulbs are needed to make turquoise cyan?
Q11B	2.21	Ν	Identify the statement about relationship between weight and resistance
			which indicates a fall at a steady rate
Q9AI	-2.05	Ν	Indicate on a diagram which of four points represents the position of the
			UK on a summer day
Q5DII	-2.07	Ν	Explain how a method of reducing power supply to a heater would work
Q3AI	-2.18	R	What does accelerating mean?
Q5A	-2.24	R	Recognise the symbol for a switch in a circuit diagram
Q1D	-2.28	R	Complete a diagram that shows some of the energy transfers, which take
			place when a radio powered a wind turbine, is switched on.
Q5DI	-2.56	Ν	If the power supply cannot be changed how can the amount of energy
			used by a heater be reduced?
Q5CI	-2.79	R	What sort of meter should be added to a circuit to measure the energy
			transfer?
Q9BII	-2.97	R	Know what a year means in terms of planetary motion
Q7A	-3.02	R	What in the Milky Way produces a hazy band of light?
Q6AII	-3.61	R	Suggest a suitable component for a light sensor
Q1C	-4.72	R	How could a family store energy for days when there is no wind for
			their turbine?

In Figure 5, the charts of the t-values for the central tier are presented. With a few exceptions, the performance of the female candidates tended to be better on the question parts that assess Sc2. For Sc3 the female performances tended to better but most of the differences were not statistically significant. The final chart is for the question parts assessing Sc4. This illustrates that the performance of the female candidates tends to be worse than that of male candidates.

(a) Paper 4 (Sc2: Life and living processes)



# (b) Paper 5 (Sc3: Materials and their properties)



Note that for two questions Q8D and Q9BII the t-values are greater than -5.

## (c) Paper 6: (Sc4: Physical processes)

# Figure 5: t-values for test of gender differences for each individual question part for the Central Tier

Table 8 contains information on those question parts for paper 4 that have significant gender differences. It is noticeable that the larger differences in favour of females relate to questions involving reproduction and medicine. However, the female candidates tended on average to perform worse than the male candidates on questions involving graphs.

Quest. ID	t-value	Туре	Synopsis
Q8AI	3.23	Ν	Identify using a diagram which birds returned after a period of regeneration after a fire
Q10EI	2.92	Ν	Given information, calculate the relative costs of transplants and dialysis
Q5AIV	2.30	R	What is the name given to the type of action where two sets of muscles work against each other?
Q9CIII	2.15	Ν	Explain why a low fertility index reduces the likelihood of fertilisation
Q9A	2.09	Ν	Complete a diagram of cattle breeding, to show the number of
			chromosomes and the number of sex chromosomes in a sperm and a egg
Q2D	2.02	Ν	Given some information, explain why an electric heart is used on the
			left hand side of the heart
Q7DII	2.01	Ν	Explain what happens to the disc from the white part of a variegated leaf when the syringe is put in the light
Q7B	-2.04	Ν	Explain what happens to glucose when it reaches the roots
Q10BIII	-2.29	Ν	Suggest why desert mice have only long loops of Henle
Q3C	-2.50	R	Suggest why the grebes give each other water weed in the dance part of their courtship
Q7C	-2.74	Ν	Explain why green leaf discs rise up and float in a syringe of water after being placed in the light
Q8CII	-4.53	Ν	Using a graph and other information, identify the year conservationists started to manage the area
Q8CI	-5.55	Ν	Using a graph and other information, identify the year in which myxomatosis started to kill rabbits in large numbers

 Table 8: Question parts with significant gender differences for Paper 4 – Central tier (Sc2: Life and its processes)

There were only a few question parts with significant gender differences in paper 5 and information about them has been presented in Table 9.

Table 9: Question parts wit	h significant gender	differences for Pa	aper 5 – Central tier	(Sc3: Materials
and their properties)				

Quest. ID	t-value	Туре	Synopsis
Q7B	3.00	Ν	Explain what it means for starch to be a polymer of glucose
Q11BII	2.86	R	What type of polymerisation is shown in these diagrams?
Q10CII	2.38	Ν	Given information about calcium and magnesium, suggest what the reaction of strontium and water would be like
Q10A	2.33	Ν	Which information in a table indicates that calcium and magnesium are metals?
Q9CI	1.97	R	Explain why solid sodium chloride will not conduct electricity, but molten sodium chloride will
Q4BI	-2.28	Ν	Given information in a table, what would the voltage be from an iron- magnesium cell?

The results of the last paper in this tier (paper 6 assessing Physical Processes) are given in Table 10. As in paper 3, there are significant gender differences in question parts about Earth and space, electricity, and mechanics. In particular, there are large differences between the sexes for some parts of a question involving Formula 1 motor racing.

Quest. ID	t-value	Туре	Synopsis
Q3CI	2.33	N	Calculate the frequency of a wave
Q4BI	-1.99	Ν	Complete a table to indicate the effect of sensors on a burglar alarm
Q6CI	-2.55	R	What causes the Earth and the planet Venus to orbit the sun?
Q8CII	-2.64	R	Explain why a perpendicular ray does not change direction
Q7E	-2.83	R	What do the ultrasound waves do to kidney stones to make them shatter?
Q8EI	-2.92	Ν	Given a circuit diagram, explain what happens to a capacitor in a flash unit when a switch is moved
Q2AII	-2.98	Ν	Explain where the energy goes when a man lifts weights
Q9A	-3.31	R	Apart from reducing drag, why is it important for the main body of a grand prix car to be close to the ground?
Q9DII	-3.35	Ν	Calculate the average speed of an accelerating car
Q9BI	-4.57	R	Suggest why a driver waves from side to side with slick tyres
Q8D	-5.11	R	Draw the path of a ray of light through a pentaprism viewfinder
Q9BII	-8.46	R	How does weaving from side to side improve the performance of slick tyres?

 Table 10: Question parts with significant gender differences for Paper 6 – Central tier (Sc4: Physical processes)

The t-values for question parts from the Further Tier are presented in Figure 6. Although the overall pattern is similar to that of the Central Tier, the magnitude of the t-values tends to be more extreme. More of the gender differences are statistically significant.



# (a) Paper 7 (Sc2: Life and living processes)



## (b) Paper 8 (Sc3: Materials and their properties)



# (c) Paper 9 (Sc4: Physical Processes) Figure 6: t-values for tests of gender differences for question parts for the Further Tier

The questions with significant gender differences for the Further tier papers are presented in Tables 11-13. This is a replication of an analysis of the previous year's examination data described in Bell, 1997a. In

particular, the results for paper 7, which assesses Sc2, are presented in Table 11. The pattern of gender differences is similar to that of paper 4 in the Central Tier.

Quest. ID	t-value	Type	Synopsis
Q7BI	4.13	R	Suggest why there is a delay after eating the glucose before glucose
			levels in the blood start to rise.
Q8B	3.28	R	Use the data given a graph to suggest why leopards have been able to
			survive in areas with continually changing conditions and the number of
			cheetahs has fallen rapidly
Q10AIII	2.90	Ν	Which cause of infertility would be most likely to be treated with 'in vitro' fortilization?
O3B	2.81	N	Fundation what happens to glucose when it reaches the roots
	2.01	N	Given information as a graph, explain which hig gat would show the
QoAIII	2.80	IN	greatest variation
Q11A	2.71	Ν	Suggest a reason why some pig farmers use a food-ball for feeding their
			pigs
Q8AII	2.61	Ν	Given information as a graph, explain why the offspring of cheetahs
			show less variation that the offspring of leopards
Q8AI	2.60	R	What is meant by the term heterozygous?
Q10BII	2.48	Ν	Explain why women treated with FSH are also given a drug to prevent
			the pituitary gland from making LH
Q7AV	2.39	Ν	Suggest two reasons why insulin has to be injected rather than be taken as tablets by mouth
Q6D	2.17	Ν	Give two effects that the removal of the tree might have on the soil
Q1DI	2.15	R	Why does a person with a joint injury have to rest?
O4BIV	2.10	Ν	Suggest two objections made by local people to the conservationist's
			methods of management
Q6B	2.07	R	Give two ways in which land might be used after clearing the forest
O10BI	2.04	Ν	Given a diagram, explain why the injection of FSH may increase the
			fertility of some women
Q1DIV	-2.06	R	Suggest why raising the injured joint above the level of the heart helps
-			reduce the swelling
Q1DII	-2.10	R	Suggest why ice reduces the swelling around an injured joint
Q9CII	-2.11	Ν	Explain the effect of rubbing sun tan oil on the amount of sweat lost
Q11DI	-2.27	Ν	Using information given in table, calculate the percentage change in the
			number of pigs produced per sow over a period of time
Q3C	-3.28	R	Explain why green leaf discs rise up and float in a syringe of water after
			being placed in the light

Table 10: Question parts with significant gender	differences for Paper 7 -	- Further tier (Sc	2: Life and
Living Processes)			

The question parts with significant differences for the chemistry related component of the Further Tier are presented in Table 12. It is noticeable that the male candidates tended on average to perform better on the question parts in which there was a degree of overlap with physics.

Quest ID	t-value	Туре	Synopsis
Q4BII	2.53	R	What type of polymerisation is shown in these diagrams?
Q3A	2.11	R	Put a ring around the name for a molecule made of a long chain of amino acids
Q1	2.08	Ν	Complete a table about metals, their uses, and the properties which make them suitable for these uses
Q6C	1.97	Ν	Use your knowledge of the electronic structure of magnesium and chlorine to explain three facts about magnesium chloride
Q7CIII	-1.95	Ν	Using information from an experiment and facts about the electrolysis of silver calculate the charge on a copper ion
Q5B	-2.12	Ν	Calculate the age of a piece of wood using carbon dating
Q8AI	-2.31	Ν	Given a table, describe one advantage of using iron rather than ceramics for cooking pots
Q2BI	-2.61	Ν	Predict the change in a glass syringe filled with air in a flask as temperature increases
Q6AI	-2.64	R	What determines the order in which the elements appear in the periodic table?

 Table 12: Question parts with significant gender differences for Paper 5 – Further tier (Sc3: Materials and their properties)

The results for the final written component are presented in Table 13. This component produced the most extreme set of gender differences for any of the components. No difference in question part performance was statistically significant in favour of female candidates. The reverse was true for seventeen question parts.

Quest. ID	t-value	Туре	Synopsis
Q9AI	1.12	Ν	Calculate the current passing through a heater
Q6D	0.97	R	Suggest one other feature of an electric drill motor which makes it more
			powerful than the model motor
Q2F	0.85	Ν	Use ideas about energy to explain why a car leaving a track at high
			speed rapidly decelerates through the gravel
Q5AIII	0.43	Ν	Use ideas about momentum to explain two answers to questions about
OTDU	0.05	N.7	the momentum of skaters
QSBII	0.25	N	Calculate the speed of one skater using the conservation of momentum
Q6A	0.22	ĸ	Draw arrows on a diagram to show the direction of the magnetic field in an electric motor
Q2DI	0.19	Ν	Calculate the gain in potential energy of a raised pendulum
Q8B	0.13	Ν	Give one use for a geosynchronous satellite
Q7D	0.01	Ν	Use energy arguments to explain the path of the rocket
Q1BII	0.00	Ν	Explain why using a water heater less would save more money than
-			using a television less
Q6EI	-2.04	R	Explain why there is an induced current in the coil of a dynamo
Q2BII	-2.15	Ν	Explain why a temperature rise is less than simple application of theory
OPCI	2.17	N	Calculate the final speed of an accelerating car
Q2CI Q7RI	-2.17	D	Mark on a diagram the direction of the force that keeps the setallite in
Q/BI	-2.22	ĸ	orbit
Q2A	-2.31	R	Apart from reducing drag, why is it important for the main body of a grand prix car to be close to the ground
O9BII	-2 31	R	Explain the difference between the energy that is transferred in both the
Q)DII	2.31	R	heater and the variable resistor
Q5AI	-2.36	Ν	How does the momentum of one skater compare with another
Q6EIV	-2.46	R	Explain why the current is always positive
Q9CII	-2.50	Ν	Given a circuit diagram, explain what happens in a flash unit when a
-			switch is moved
Q7BII	-2.64	R	What provides the force to keep the satellite in orbit?
Q5C	-2.67	Ν	Calculate the kinetic energy of a skater
Q8D	-2.93	R	Give a reason why particular sound waves are transmitted more
			effectively by tweeter than by woofer
Q9CI	-3.19	Ν	Given a circuit diagram, explain what happens to a capacitor in a flash unit when a switch is moved
O6CII	-3.23	R	Explain why many sets of rotating coils make a motor more powerful
	-3 30	N	Calculate the average speed of an accelerating car
08EI	-3 64	R	Explain why the sound is received more strongly in all directions from
YOLI	5.04	K	the tweeter than from the woofer
Q7C	-3.86	R	Give two reasons why cannon balls fall to earth but satellites do not

 Table 13: Question parts with significant gender differences for Paper 9 – Further tier (Sc4: Physical processes)

This section has demonstrated that there are significant gender differences for the retrieval question parts. In the next section, the processes involved in answering retrieval questions will be considered. This is done in an attempt to provide an explanation for these differences.

## A model for the processes required to answer a retrieval question

In this section of the paper, the processes required to answer a retrieval question will be considered. By defining a model that considers the various stages involved in answering such question parts, it is possible

to develop explanations for differences in performance in retrieval questions. In particular, these explanations will be used to investigate the various hypotheses that have been used to account for the gender differences found in science performance.

On the quiz show, *Who wants to be a millionaire*?, Chris Tarrant said to the audience that 'A question is only easy if you know the answer, otherwise it's impossible'. This seems to be the obvious explanation of difficulty for a question which involves recalling factual information (for the purposes of this paper, this will also include questions involving recognising the correct answer even though they are cognitively different processes).

This explanation is clearly unsatisfactory for examination questions. In theory, and it is hoped in practice, candidates will have been exposed to all the information in an examination. In spite of this the facilities of questions solely involving recall and recognition vary. In this paper, research into memory will be reviewed to identify potential features of questions that could account for these differences. Wilding et al. (1999) noted that the relationship between other cognitive abilities and examination performance has not been intensively investigated, apart from the obvious case of general intelligence.

To answer a retrieval question, a number of stages that can be identified. A simple model used to investigate the demand of retrieval questions is presented in Figure 2. A more complicated model for answering questions is given in Pollitt and Ahmed (1999)



## Figure 7: A simple model for a retrieval question

This model is a simplification of the actual processes involved (for example, short-term memory has been included in working memory). In this model six processes have been identified. These are

- 1. Exposure/observation The exposure of the information to the candidate.
- 2. Encoding The formation of a long-term memory and the creation of a network.
- 3. Query Interpreting the question
- 4. Interrogation Send a query into long term memory
- 5. Retrieval Recovering the answer for from long term memory
- 6. Expression The conversion of the information into a suitable response for the examiner.

For a candidate to answer a question successfully, all six processes have to be completed successfully. Obviously the first two processes (denoted by solid lines) can occur over a considerable period before the actual examination and the last four processes (denoted by dotted lines) occur during the examination. The candidate will be unable to answer the question if something goes wrong with at least one of the above processes. Firstly, the candidate will not be able to answer the question if there is no long-term memory to recall. Question difficulty varies if some information is more difficult to encode than other information is. Thirdly, if the candidate does not understand what is required by the question they will be unable to answer the question. The cues to the answer may not seem relevant to the candidate. Finally, the candidate may not be able to express the memory in a suitable form (e.g. the candidate might be able to visualise a bicycle but be unable to describe it in words). In the remainder of this paper, the properties of each of these processes that influence the demand of the question will be considered.

By considering each of these features in turn, it is possible to identify potential problems, which might prevent the successful response to a recall question. If these difficulties could be rated then it might be possible to predict the difficulty of the questions.

Exposure is the most trivial part of the process. For a fair examination, it is necessary that the information that the candidate has to encode (i.e. the facts to be learned) will have been specified in the syllabus. It can be assumed from this that the candidate has had the opportunity to create memories about information. Obviously, if there has been no exposure to the information the question cannot be answered by recall. Candidates will fail at this stage if for some reason they miss out part of the syllabus and do not try to rectify this problem. At the candidate level, this would imply that there is a lack of motivation or opportunity. A gender difference in performance arising at this stage of the process would arise as a result of the difference in attitudes towards science and learning in general.

It is not enough to have been exposed to the information by sensory stimula. Encoding is necessary. Although in the figure encoding is represented as a one-way process from working memory to long-term memory, this is obviously a simplification. When the new information is in working memory, old information is extracted from long-term memory and is related to it. This is a vital part of the process. Previous experience is essential in interpreting new information. For example, a surgeon has to be trained to recognise that a shadow on an X-ray may be a serious tumour. Encoding depends on previous knowledge, which is necessary to encode it. The classic example of this relates to chess players. de Groot (1965) and Chase and Simon (1973) showed that master chess players can reconstruct far more than novices of a briefly glimpsed chess position from the middle of the game, so long as the pieces are in realistic settings. With random positioning of the pieces there is no superiority in the chess master's recall.

This leads to the question of why information is not encoded. People do not memorise an exact copy of the stimulus. They may look only at certain features of the stimulus. Therefore, it is necessary to consider what individuals think about when they are presented with something to be remembered. It cannot be assumed that they are trying to remember the whole thing. In models of memory, a distinction is made between the encoding (forming a memory) and storage (maintaining a memory). Craik and Lockhart (1972) proposed that strength of memory depends on how deeply information is processed, not on how long it is processed. There is experimental support for this: memory for words was not improved by merely repeating them for a longer period of time (Glenberg et al., 1977).

Encoding is more likely to be successful if the information is structured and if it links to many aspects of an individual's experience. There are two ways that information can be structured: real and artificial. Mnemonics are examples of such structures. It can be argued that relying on artificial structures is not as desirable as understanding the real structures.

At this stage of the process, there are several factors that can inhibit performance on the questions. Firstly, lack of interest by the candidates can mean that even repeated exposure to material does not lead to successful encoding. Secondly, successful observation requires that the candidates have sufficient previous experience to develop the theories necessary to understand what is being observed. A gender difference in performance would arise if there were prior differences in experience between male and female candidates even if they received the same teaching experience in science lessons.

The first step in answering a question is to bring into working memory the question and use the content of the question to generate a set of cues, which are used to start a search of long-term memory (described as reading in Pollitt and Ahmed (1999)). At this point, a failure to understand the question could lead to the generation of the wrong set of cues or stop the question answering process completely. Part of this process is deciding to discard irrelevant information. Question context could have an effect here. It should be recognised that cues are knowledge and form part of the network of related knowledge that includes the answer. Answering the question involves finding a linkage (there may be more than one) between the information that forms the set of cues and the information that forms the answer.

Note that usually the query process and the next two steps, the interrogation and retrieval processes, will tend to form an iterative loop. An initial set of cues is generated for use in a search of long term memory. The results of the retrieval processes are used to generate more cues if the answer is not found.

In the encoding process, information has been logged in long-term memory and has been categorised or classified in an appropriate way. The way of recalling the memory is to present all or part of the code that was laid down when it was classified. Using the library analogy, if the play Hamlet has been classified under "Shakespeare", "Drama", "Verse", "Prince" and "Denmark", then some subset of these cues would be enough locate it. These retrieval cues are snippets of information that allow one to access a memory trace.

There has been a considerable amount of research into this process. Bahrick (1984) tested college graduates to see how much they remembered of Spanish language classes that they had taken in high school or college. Most of them had made no use of the language after finishing the course, so this was recall without intervening practice. Retention seemed to have levelled out over very long periods of time. Bahrick suggested that some memory is retained in a 'permastore' which is unaffected by the passage of time. Those students who had learned better originally remembered better over the years.

More recently, Cohen, Conway and Stanhope (1992) examined the very long term retention of knowledge which was originally acquired from formal education and which had not been subsequently used or reviewed. Three different knowledge domains were investigated, cognitive psychology, experimental design and statistics, and literature. In each of these subject areas between 300-450 former students of the Open University completed a series of memory tasks designed to test retention of the course material. The general pattern of the results showed that rapid forgetting occurs in the first two years after learning. Thereafter memory stabilised and remained at above chance levels for the remainder of the period. Knowledge that which is still retained after approximately two years appears to remain intact indefinitely. There were two exceptions to this pattern of forgetting. Although memory for details and highly specific facts declined rapidly, memory for general principles was extremely stable and showed no forgetting over the 12 year period. Similarly, knowledge of experimental design and statistics also showed no significant decline over the whole of the 12 year retention period, suggesting that active use of this knowledge in practical work made it more memorable, and possibly reflecting a superior retention of procedural, as opposed to declarative, knowledge. This would suggest that, for diligent candidates, questions involving procedures will be more likely to be encoded than questions involving memory retrieval. The results of this research are not surprising because the solving of problems that is involved in learning procedural skills inevitably leads to a greater depth of processing.

Although a candidate may be able to retrieve the knowledge, it is possible that the candidate would be unable to express the answer in a way that would satisfy the examiner. A simple example of this would be a question that requires a candidate to complete a diagram. The candidate might be able to visualise the solution in their mind but not be able to reproduce the image on paper. More subtlety, this process includes the effect of response format. For example, requiring extended writing for an answer can also require additional linguistics skills in addition to retrieval.

## Discussion

In this paper, it has been found that gender differences exist in question parts that only involve the retrieval of declarative knowledge and not the use of procedural knowledge. A consideration of the processes used would suggest that the likely causes for each difference relate to differences in attitude and prior experience. These must reflect an interaction between biologically based differences in the types of activities that boys and girls prefer to engage in and the environments made available to them by parents and by peers. The performance difference described in this paper has plausibly been attributed to the very different leisure activities of young boys and girls out of school (Johnson and Murphy, 1986; Johnson, 1987).

It should be noted that differences in experience may develop very early. Servin, Bohlin and Berlin (1999) used a cross-sectional study to investigate toy-choice in 38 one-year-old, 33 three-year-old, and 35 five-year-old children. They could choose between 10 different toys (four feminine, four masculine, and two

neutral) in a structured play-session. Girls and boys chose different toys at all three ages. *In another* study, it was found that people were the focus of 80% of the stories of 2-year-old girls, relative to 10% of the boy's stories; boys primarily talked about objects, such as cars and trains. By 4- years-of-age, 100% of the girls and about 60% of the boys talked about people, as opposed to objects (Goodenough 1957).

It could be argued that retrieval questions are about assessing the networks of knowledge. A *Retrieving scientific information* question is assessing whether the candidate's network of scientific information is able to link the information in the question with the information required by the answer. In research into memory, it has been suggested that it not that people forget but that they are unable to find memories. This has implications for the concept of fairness in question design. Some authors have argued that questions can be unfair if candidates know the information but are unable to recall it in a certain context. If it is the network that is being assessed then the issue is whether the question is assessing a realistic, important and reasonable linkage between knowledge given as cues in the question and knowledge required as an answer. Given the existence of differences in attitude and early experience, it is predetermined that males will outperform females in certain parts of science, and vice versa. However, one must consider that there are some curricular areas, such as electrical circuits or earth/space topics, where it would be impossible to construct items that are not in a male sphere experience.

## References

ARCHENHOLD, W.F., BELL, J., DONNELLY, J., JOHNSON, S. and WELFORD, G. (1988) *Science at Age 15: A review of APU findings 1980-1984.* London: HMSO.

BADDELEY, A.D. (1999) Essentials of Human Memory. Hove: Psychology Press Ltd.

BAHRICK, H.P (1984) Semantic memory content in permastore: fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*. 113, 1, 1-29.

BELL, J.F. (1997a) *Sex differences in Performance in Double Award Science GCSE*. Paper presented at the British Educational Research Association Annual Conference, University of York, 1997, 10-14 September.

BELL, J.F. (1997b). Evidence for the continued existence of the Physics 'gap' at age 16. *British Journal of Curriculum and Assessment*, 7, 35-39.

BELL, J.F. (1998) *Patterns of subject uptake and examination entry 1984-1997*. Paper presented at the British Educational Research Association Annual Conference, Queen's University, Belfast, 26-30 August. BELL, J.F. (In prep.) Science subject uptake for year 11 pupils (1974-1997)

CHASE, W.G., and SIMON, H.A. (1973) Perception in chess. Cognitive psychology, 4, 55-81.

COHEN, G., CONWAY, M., and STANHOPE, N. (1992) Very long term retention of knowledge. HCRL Technical Report No. 87. Milton Keynes: Open University.

CRAIK, F.I.M., and LOCKHART, R.S. (1972) Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671-684.

DE GROOT, A.D. (1965) Though and choice in Chess. The Hague: Mouton.

GLENBERG, A. M., SMITH, S. M., & GREEN, C. (1977). Type I rehearsal: Maintenance and more. *Journal of Verbal Learning and Verbal Behavior*, 16, 339-352.

GOODENOUGH, E. W. (1957) Interest in persons as an aspect of sex differences in the early years. *Genetic Psychology Monographs*, 55, 287-323.

JOHNSON, S. (1987) Early developed sex differences in mathematics and science in the UK, *Journal of Early Adolescence*, 7, 21-33.

JOHNSON, S. and MURPHY, P. (1986) *Girls and Physics. Reflections on APU survey findings.* APU Occasional Paper No. 4. London: Assessment of Performance Unit.

POLLITT, A., and AHMED, A. (1999) *A new model of the question answering process*. Presented at the DAEA Conference. Slovenia.

SERVIN, A., BOHLIN, G., and BERLIN, L. (1999) Sex differences in 1-, 3-, and 5-year-olds' toy-choice in a structured play-session. *Scandinavian Journal of Psychology*, 40, 1, 43-48.

WILDING, J., VALENTINE, E., MARSHALL, P., and COOK, S. (1999) Memory, IQ and examination performance. *Educational Psychology*, 19, 2, 117-132.

ZEITOUN, H.H. (1989) The relationship between abstract concept achievement and prior knowledge, formal reasoning ability and gender. *International Journal of Science Education.* 11, 2, 227-234.