서울大學校 師大論叢 第33輯(1986. 12. 30)

Some Considerations on Making Questions

—After seeing the materials of assessment of performance unit of England—

Hui Soo An
(Department of Earth Science)

Introduction

The assessment by means of written questions is frequently used in order to obtain informations about a population. It is necessary to consider several stages to achieve the aim of the exercise. Firstly, design appropriate questions, secondly, measure the phenomenon using them, thirdly, analyse the results and finally, evaluate the population and consider the feedback. All of the four stages are complex. In the first stage, it is necessary to identify the population to be measured and the types of questions suitable for it. In the second stage, the sample sizes have to be chosen in order to allocate the samples and questions. In the third stage, the mark scheme is constructed and evaluated. In the fourth stage, all previous stages are evaluated and the feedback is considered. The first and third stages are related directely with producing questions, and at these stages, the experiences of educational experts are needed.

Fortunately I have had a good opportunity to see and investigate some of the question materials and results in APU (Assessment of Performance Unit). The APU was developed in the United Kingdom in 1980 in order to survey childrens' scientific performance. As reported by the APU (1983, 1984a, b, c, 1985) the surveys were carried out on a random samples of 11 years old pupils in primary and middle schools and of 13 and 15 years old pupils in middle and secondary schools. The samples were collected from random samples of schools through the whole area. Pupils were drawn at random from each school giving a total sample of between one or two percent of the population at each age.

The questions have been published in the APU science books by Department of Education and Science (see the previous references).

The range involved in the assessment of performance was grouped into the six categories as shown in Table 1 (APU, 1984).

Table 1. The Categories of Science Performance

1.	Use of graphical and symbolic representation	—reading information from graphs, tables and charts	written test
		—representing information as graphs, tab les and charts	
2.	Use of apparatus and measuring instruments	—using measuring instruments	group practical test
		—estimating physical quantities	
		—following instructions for practical work	
3.	Observation	—making and interpreting observations	group practical test
4.	Interpretation and application	—Ⅰ. interpreting presented information	written test
		—Ⅱ. applying: Biology concepts	
		Physics concepts	
		Chemistry concepts	
5.	Planning of investigations	—planning parts of investigations	written test
		—planning entire investigations	
6.	Performance of investigations	—performing entire investigations	individual practical test

In this study eleven questions, which had relatively high rate of response, were investigated on the point of view what is the problem appeared in real assessment.

As quoted by the APU (1985), "...thus the results of individuals have little meaning until added together to give a profile for the population at a given age.", there is little meaning in a particular question unless they are summarized in a large category.

The purpose of this study is not to assess the level of performance of each question, but to investigate general features of them.

Data and check points

The data used in this study are shown in Table 2.

They are chosen because they mostly have a high rate of response. Some of them have relatively high average scores around 50% whereas some are very low. The high rate of response means that they drawn out much information from the pupils even though their responses were not correct.

Table 2 shows the type of test, mapped mark and standard deviation of mapped mark. The mapped mark is the standardized scoring system scaled from zero to three, which requires no further explanation.

The type of test is specified into written and practical tests.

50 examples of each question were investigated. They were scored on the basis of the

Table 2. The data used in this study.

No.	Name	Category	Type	Rate of Response	Mean	Mapped mark	Standard deviation
B1316	ores	3	pr	100%	45%	1.40	0.80
B1348	coal	3	pr	99%	51%	1.50	0.60
B1388	ore key	3	pr	92%	47%	1.40	1.00
E1496	iced soup	4	wr	97%	27%	0.81	0.94
B1566	coolers	4	wr	90%	3%	0.09	0.32
B3491	mineral similarities	3	pr	—	—	—	—
B4071	draining cups	4	wr	95%	3%	0.08	0.36
B4077	freezer bag	4	wr	91%	21%	0.62	0.83
B4163	jack-in-the-box	4	wr	98%	56%	1.68	0.96
B4168	pebble plus	4	wr	97%	49%	1.47	0.95
B4209	cold sense	4	wr	97%	5%	0.16	0.41

wr=written test and pr=practical test.

original mark scheme at first and then rescored based on a new reference in which every possible reason had been taken into consideration.

The following points were checked.

1. To check whether there are any difficult expressions in the question.
2. To check whether the mark scheme was made reasonably in relation to question.
3. To check whether the tester who structurize question considered the way pupils think.
4. To check how much the tester took into consideration the particular words that pupils use.
5. To check whether the foci of the question coincide with the number of key words in the mark schemes.

Analysis

The following are the individual analysis of questions (See the appendices).

(1) B-1316 is a question to assess the pupils' ability to carry out a series of actions and observe the consequent results.

The question itself is structured very nicely in order to see how pupils observe streak colour, the response of the rocks to the acid and the description of the rocks.

But when we see the results of the pupils' response, a problem could be found in question and mark scheme.

In part (a) of the question would be better if pupils had already learned about streak

colour and how to scratch minerals on a tile to make the streak colour, (actually it is rather difficult to make a streak colour even to one who knows what streak colour is). The tile used in the test is too weak to make streak colour and some pupils failed to streak.

In part (b) of the instructions "Write down below what happens to each rock" would be better to be replaced by "Observe whether the rocks fizzle or not". Many pupils answered to above question for iron pyrite "Its colour changed to yellow" and "Its colour changed to black" to haematite instead "no response" to them.

In part (c) of the instructions "briefly describe its appearance" implies the use of single term or sentence, but the mark scheme required more than two answers.

Here, two results can be compared; one is strictly based on the scheme and the other takes into consideration the various possibilities which are described above.

Figure 1 shows the average scores and ranges in both results.

In the latter case the mean score moved a little upwards and the range increased.

There is no special significance in the increase of mean score and range. It is simply because the correct evaluation of right result is important.

(2) B-1348 is a question which tested the pupils ability of observation. The presented materials were coal (labelled P) and chalk (labelled Q).

The mark scheme required ten pairs of comparison words describing the features of rock and chalk.

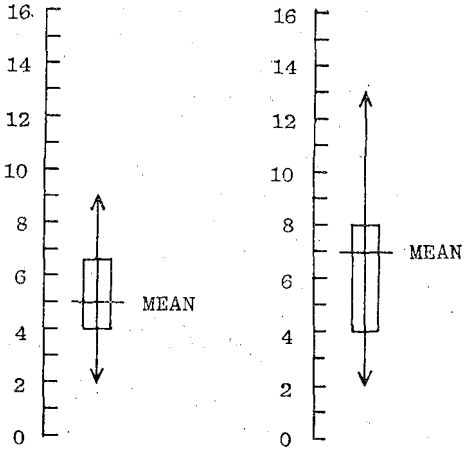


Fig. 1. The average scores and the ranges of scores on the original mark scheme (left) and on the other takes into consideration the various possibilities.

Table 3 shows the descriptive words which pupils had written in their descriptions. 49 such words were used and the total frequency is 443.

White and black are most frequent ones. We notice that most of the description are related to the appearance, feeling and figure. An interesting one is that the frequencies of "not" and "than" were quite high. "not" is used in negative comparison like black and not black and "than" in comparative comparison like "rough" and "more rough than", and they all were fully marked.

Taking into account the previous examples, the mark scheme is unreasonable. Counting the number of adjectives which appeared in pupils' answers seems to be reasonable because most of the pupils lined adjectives without thinking its comparison.

When we make a mark scheme, it is very important to consider the type of words that pupils use. The tester is apt to construct a mark scheme according to adult's opinions, not children's.

Figure 2 shows the range of marks and the average scores in both cases, one is based on the original mark scheme and the other on the new scheme in which scores represented the number of adjectives.

(3) B-1388 is a question in which pupils have to use a key to identify objects.

46 pupils out of 50 gave the correct answer to slate, 19 pupils to granite and 18 to

Table 3. The descriptive words which pupils used in their descriptions.

Figure		Superficial appearance		Feeling		Inference		Relative		
layer	16	white	43	hard	17	chalk	19	not	38	
sharf edge	13	back	42	crumble	17	coal	17	than	20	
smooth edge	10	shiny	19	smooth	16	solid	4	different	9	
crystal	6	rough	18	soft	15	pure	1			
piece	6	dusty	12	light	10	combination	1			
round shape	5	dull	9	heavy	10					
chip	5	dark	6	powder	6					
big	2	glitter	2	stain	4					
irregurar	2	clean	2	ragged	3					
small	1	tranparent	1	black dust	2					
bumpy	1	bright	1	white	2					
grain	1	dry	1	fragile	2					
flake	1			brittle	2					
dinty	1			jaggy	1					
				smell	1					
14	70	12	156	15	108	5	42	3	67	443

haematite. As shown in Table 4, the number of steps which need to find out correct answer of slate, haematite and granite are two, three and four respectively.

Is there any relationship between frequency and steps? In order to consider this relationship it is necessary to investigate the distribution of wrong answers.

Table 5 shows the vocabularies appeared as wrong answer and their frequencies.

Table 4. The frequency and steps of correct answer.

correct answer		frequency	steps
haematite	(R)	18	3
slate	(S)	46	2
granite	(T)	19	4

Table 5. The vocabularies appeared as wrong answer.

correct answer	R (haematite)		S (slate)		T (granite)	
vocabularies in wrong answer	mica	11	chalk	1	mica	7
	limestone	4	haematite	1	sandstone	3
	sandstone	1			iron-pyrite	6
	iron-pyrite	2			haematite	9
	granite	10				

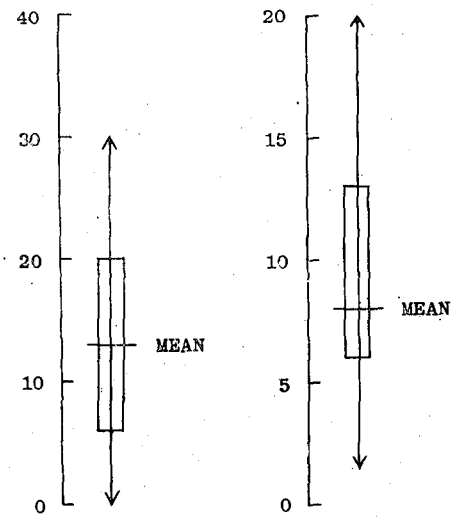


Fig. 2. The average scores and the ranges of the scores in both cases, one is based on the original mark scheme (left) and the other on the new scheme in which scores represented the number of adjectives.

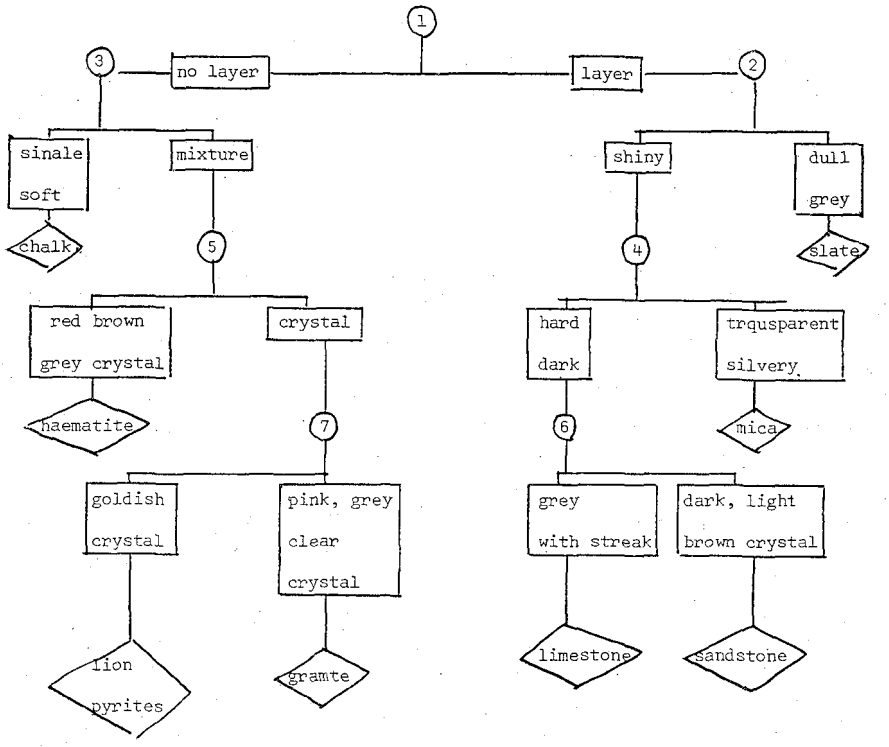


Fig. 3. The flow diagram accordingto the clues of the questions.

The uppermost minerals are correct answers, and the minerals below are the wrong answers and their frequencies. Comparing this table with the flow diagram (Figure 3), the first clue is to decide whether the mineral is layered or not.

If they chose anyone among chalk, haematite, iron-pyrite, and granite where slate was correct answer, or if they chose anyone among slate, mica, limestone and sandstone where haematite or granite was a correct answer, it means that they did not know the meaning of layer. According to Table 5, exactly half of the wrong answers come from the misunderstanding of the meaning of layer, that is, the pupils were likely to think of layer as a flat surface.

The wrong answers of iron pyrites and granite instead of haematite and the wrong answer of haematite instead of granite seem to come from a misunderstanding of clue 5 in Figure 3. The wrong answer of iron-pyrites instead of granite is also likely to come from the vagueness of clue 7.

What can be said about this question is that pupils seem to be hindered by the complex step needed to trace the correct answer. The wording of the clues should be clear so that

pupils are not bewildered when making their decisions.

(4) B-1496 is a question in which pupils have to think and explain the reasons scientifically.

Part (b) in the mark scheme is a typical example, which shows the difference in the way of thinking between tester and pupils. In this mark scheme the tester referred to the air spaces inside cells of the vegetables and fruit but most of the pupils referred to the amount of water inside. Anyway, the explanations are all reasonable and scientifically correct.

Table 6 shows the distribution of scores based on the mark scheme (old) and on the revised scheme (new).

Table 6. The score distributions in B-1496.

score	0	1	2	3	4	5
old	36	3	10	1		
new	36	3	3	1		7

In any case 36 pupils out of 50 gave no response. The rate of respose is also very low. 10 pupils scored two in the old mark scheme but if we remarked them with the new scheme, seven of the ten moved to the highest mark.

This means that the old mark scheme could neglect many of the pupil's opinions.

(5) B-1566 assesses the ability to think scientifically. This is one of the questions whose rate of response is quite low. None of them could explain why the water in the earthenwear jar keep cool.

The frequent explanations were that the water passes through the holes. Although the exact meaning is not clear, they tried to relate cooled water and holes. Only one pupil used the term "evaporate" but his explanation was not correct.

Why is the rate of correct answer so low compared with other questions? This is probably because the subject treated here is an unusual thing which they could not experience in daily life.

(6) Question B-3491 is to assess pupils ability to identify difference between two minerals. One is muscovite and the other is black slate.

In this question there are some problem to consider.

The first is found in the sentence of part (a), that is, "Write down five ways in which the minerals are different. Take each difference in turn and describe the difference for each mineral". In this sentence there are two requests; "Write down five ways" and "describe the differences".

Seven pupils did not respond in the difference but in terms of ways. Mistakes like this are found in part (b) as well.

The second problem is found in the mark scheme which give four marks for a complete and correct comparison and one mark for a non-comparative answer. The interpretation of "comparison" seems to be somewhat different between tester and pupils, for instance, pupils seem to think that "black and shine" is a comparative pair but commonly they think black and white as a comparison with strict constrainment.

Table 7 shows the distribution of marks. The mean of the new mark distribution is increased compared with that based on the old mark scheme.

(7) In question B-4071, pupils tend to misunderstand the meaning because the point of question is not explicit. There are some obscure expressions so that whoever read this paragragh probably imagine of some other causes sticking cups and glasses with the surface.

11 pupils out of 50 shows a misunderstanding in their answers. They thought that something in the hot water or hot air makes cups or glasses stick to the tray as they cooled.

"What happens to the air in the glass when the glass is first put down on the surface (remember that the washing-up water was hot)." This question is also likely to be misunderstood by pupils because the previous explanation has vagueness in itself. The evidence for this is that not only do 19 pupils out of 50 respond to this question incorrect in terms of the trapping of air inside the cups, but also they are unable to give any further explanation. All but four of the pupils were unable to explain or understand the reason why the cups stuck to the surface of the tray.

28 pupils gave a totally incorrect response.

In the mark scheme there is bad consideration. The pupil's way of thinking seems to be very simple. When they were asked a single question they responded with a single answer and when asked a plural question, their responses are in plural. In this question the number of questions does not match with those of the answers expected by tester. There is a problem in the figure presented with the question. The small black spots around the cups are likely to be a cause of misunderstanding. They may be considered to be sticky materials.

Table 7. The distribution of mark in B-3491.

mark	0~5	6~10	11~15	16~20	21~26
old	9	18	14	7	2
new	3	7	7	18	15

Table 8. The distribution of mark in B-4077.

mark	0	1	2	3
old	36	4	8	2
new	21	0	0	29

Pupils definitely seem to dislike writing long sentences and have little ability to explain complex reasons step by step.

(8) B-4077 is also a question requiring an explanation in a long sentence. The mismatching between the number of questions and answers is a common problem as seen in previous questions. Table 8 shows the distribution of marks of old and new mark scheme as previously.

The distributions are quite different between the old and new mark scheme. In the new distribution, 29 pupils obtained 3 marks and the rest of 21 pupils zero marks. The reason is very simple because they responded to this question with a simple sentence as shown following example, "The pressure of water on the back squeezes the air out." Only two pupils wrote an absolutely correct answer.

There are also some problems in the old mark scheme. Table 9 shows the words and their frequencies that pupils used in their answers.

Table 9. The words and frequency.

words	frequency
remove	16
Push out	11
squeeze out	10
pressure	9
force out	4
heavier	2
compress	1

The most frequently used word is "remove" and the next "push" and "squeeze out". These word are different from those referred to in the mark scheme. The tester should consider children's vocabularies and their ways of thinking before completing mark schemes and questions. The mean score of 0.52 based on the mark scheme is a little lower than the total sample mean. But if it is assumed that the words shown previously were correct ones, the mean score would be changed abruptly to 1.74.

(9) Question (B4163) would be a good one if a little more consideration had been given

Table 10. The distribution of mark in B-4163.

score	0	1			2			3
code	D	C	E	H	B	F	G	A
freq.		16	1				33	

to the mark scheme.

Table 10 shows the distribution of marks based on the original mark scheme. There are no pupils who scored zero or three. 33 pupile scored two marks and the remaining seventeen pupils one mark. This mark distribution seems to be completely affected by the mark scheme. If you refer the question, "Where is the energy before the box is opened, and how did it get there ?" The two underlined parts are the main points of the question.

Most of the pupils responded as if the question required two answers. In Table 10, 33 pupils (66%) gave code G answer, which means that they answered like this "The energy is in the spring and it is stored in it when the spring is pushed down". If the tester wished pupils to answer as expected in mark scheme (ii) another question should be asked: "From where is the energy transferd to the spring".

In the mark scheme (iii) the term "compress" is badly chosen. As shown in Table 11, 21 pupils used "push down" whereas only three pupils wrote "compress".

Table 11. The words and frequency.

words	frequency	words	frequency
push down	21	pull	1
compress	3	squash	1
put	2	undo	1
press down	2	force down	1

Table 12. The frequency distribution.

score	0	1	2	3
frequency	8	20	0	22

(10) The rate of response in question (B-4168) is 96 percent. This subject is quite suitable for the pupils because the context of the problem is a common thing which everybody might experience in their daily life.

As shown in Table 12, there are no pupils who scored two marks. Nine pupils out of 20 who scored one mark are likely to know the principle of this concept although they did

not explain it explicitly. As a result more than half of them seemed to know the correct answers.

(11) As in question B-4168, there are no mistakes in question B-4209. 15 pupils out of 50 used the word "conduct" and "insulate" but they seemed to hardly understand their meanings. The response rate is 100 percent but 70 percent of them wrote the wrong answer. Most pupils seemed to dislike and were unable to explain a little long sentence.

Results

I analysed and investigated twelve questions and their mark schemes. Many of them contain ambiguities and they seemed to affect the results. Table 13 shows the summarized informations.

Table 13. The summarized information.

No. and name	problems with the question	mark scheme
B-1316 ores	prerequisite learning is needed and ask whether or not	1 mark to 1 answer total 16 to 12
B-1348 coal	complicate and difficult	mark scheme counting the descriptive words used
B-1388 ore key	suitable	suitable
B-1496 iced soup	none	mark scheme change to single checked problem
B-1566 coolers	subject is unusual	none
B-3491 mineral similarities	expression is ambiguous	mark scheme counting the descriptive words used
B-4071 draining cups	context is misundee standablr	mark scheme is constructed too specifically
B-4077 freezer bags	none	mark scheme check in single and counting words used
B-4163 jack-in-the-box	none	part (ii) of the mark scheme is unnecessary
B-4168 pebble plus	suitable	suitable
B-4209 cold sense	a little difficult	none

Conclusion

It is not possible to draw general conclusions from only eleven questions. It is very important opportunity to see actually the questions and the APU results. The purpose of this study is not to assess the APU itself.

Thus the following conclusions are made.

1. The subject which is chosen as the main component of question should be something which pupils come across in their daily lives.

2. The number of key points in a question should coincide with mark scheme, especially the allocation of marks.

3. The tester must know the kind of words that pupils tend to use in their expressions.

4. The rates of correct answer are very low in the questions which require complex thinking steps or a long explanation. In those cases, the context of question is rather to divide into several sub-questions.

5. It is important that the question and mark scheme should be checked repeatedly before the assessment. This is because children's perception of the question is sometimes very different from that of the tester.

By considering all the points above we could draw much information from pupils, regardless of whether their responses were correct or not.

Acknowledgement

The author is indebted to Professor Paul Black of the University of London who allowed him to use the materials of the assessment of performance unit.

References

Department of Education and Science, Welsh Office, Department of Education for Northern Ireland (1983), Science at Age 11, Science Report for Teachers: 1, p32.

———(1984a), Assessment framework Age 13 and 15, Science Report for Teachers: 2, p48.

———(1984b), Science at Age 13, Science Report for Teachers: 3, p44.

———(1984c), Assessment framework Age 11, Science Report for Teachers: 4, p44.

———(1985), Practical test at Ages 11, 13 and 15, Science Report for Teachers: 6, p54.

〈국문요약〉

문항출제시에 고려될 사항에 관한 연구

안 희 수

영국에서 개발된 과목성취도 평가법(APU: Assessment of Performance Unit)에서 사용하고 있는 문제은행중에서 몇개의 사례를 선택하여 성취도 평가문제를 작성 할때에 나따날 수 있는 문제점을 검토하였다. 선택된 문제는 11문제이고 그중에는 실기문제테스트가 4문제, 실기시험 문제가 7문제였다.

나타난 문제점으로는 첫째, 문제의 주제가 되는 것이 일상적 생활범주에서 선택되지 않았을 경우에는 질문에 대한 응답율이 낮을 뿐만아니라 정답비율도 낮아진다. 둘째, 주관식 문제를 채점함에 있어서 사용하는 정답모형(모범답안)은 질문의 요점 및 질문수와 일치하여야 한다. 셋째, 모범답안을 작성시에는 평가집단이 사용하는 표현어휘를 참조하여야 한다. 넷째, 복잡한 사고나 설명이 필요한 문제를 질문할 경우에는 몇개로 나누어 질문하는 것이 좋다. 다섯째, 질문문제와 정답모형이 완성되었을 때라도 몇번의 예비평가 작업을 통하여 교정하여야 한다.

Appendices

(1) Question B-1316

You have been given large lumps and some small pieces of four different rocks, each labelled with its name, a piece of tile and some acid in a dropping bottle.

a) Take a small piece of each rock and make a scratch with it on the tile. You may use both sides of the tile. Write down the colour of each scratch:

b) Now take the small piece of each rock you have used before. Place them one at a time on your tile and add a drop of acid. Write down below what happens to each rock.

c) Take a large lump of rock in turn and *briefly* describe its appearance:

Malachite .....
Iron pyrites .....
Haematite .....
Limestone .....

mark scheme

Colour of scratch—	*Codes*—no code applies if all correct,		
1 mark per correct answer	or all incorrect		
M—Green	L —incorrect—A	4	
IP—Grey/Green	L,H —incorrect—B		
H—Red/Brown	L,H,IP —incorrect—C		
L—Colourless/White	Other combinations—D		
Reaction with acid—	*Codes*—no code if all		
1 mark per correct answer	correct or incorrect		
M—Fizzes/Changes colour	IP,H —incorrect—A	4	
IP—No reaction	IP —incorrect—B		
H—No reaction	H —incorrect—C		
L—Fizzes	Any combination		
2 marks per mineral—from	but L *or* M —incorrect—D		
Malachite —different shades of green	Non of these—incorrect—E		
—in layers/distinct pattern			
—hard		8	
—smooth			
Iron pyrites—shiny bits of gold			
—grey-black			
—hard	*Codes*—for C—attributes mentioned		
Haematite —red/brown and grey	colours only —A		
—shiny bits	textural features only —B		
—hard	hardness only —C		
—bumpy	—combinations—		
Limestone —grey-white	colour & texture —D		
—rough	colour & hardness —E		
—hard	textue & hardness —F		
—crystals/shiny	all 3 —G		⑯

(2) Question B-1348

You have been given two different pieces of rock which have been labelled P and Q.

Look very carefully at the two rocks. Your job is to find as many differences as you can between the two pieces.

Take each difference in turn and describe it for each rock.

P	Q
.....	.....
.....	.....
.....	.....
.....	.....

mark scheme

Score up to 10 comparison awarding		
4 marks for a complete and correct *comparison*	40	
e.g. P is black, Q is white		
OR		
If no comparison is made between the two rock samples,		
Score 1 mark for each correct observation	20	
e.g. P is shiny, Q is soft		㊵
Record score under two headings—		
A—Score for comparative statements		
B—Score for single specific statements		
Codes—use three columns for major attributes used in answer		
C—colour		
D—texture/grain		
E—hardness		

(3) Question B-1388

You have been given three samples of different rocks labelled R, S and T. Try to find the names of each rock in turn by working through the sets of 'clues' below. Always start at clue number 1 for each rock.

Write the name of the rock next to the label at the bottom of the page.

1. The rock is formed in layers	If YES go to 2.
The rock is not formed in layers	If YES go to 3.
2. The rock is shiny in places	If YES go to 4.
The rock is dull and grey	SLATE
3. The rock seems to be a mixture of different substances	If YES go to 5.
The rock seems to be a single substance which is soft and powdery	CHALK
4. The rock is transparent in places, and silvery and brittle	MICA
The rock is hard and dark in colour	If YES go to 6.
5. The rock is red-brown all over with patches of small, shiny grey crystals	

on the surface	HAEMATITE
The rock has crystals right through	If YES go to. 7.
6. The rock is dark brown and light brown	SANDSTONE
The rock is grey with streaks of white through it	LIMESTONE
7. The rock has gold coloured crystals over its surface	IRON PYRITES
The rock has pink, grey and clear crystalline patches	GRANITE

Rock R is called.....

Rock S is called .....

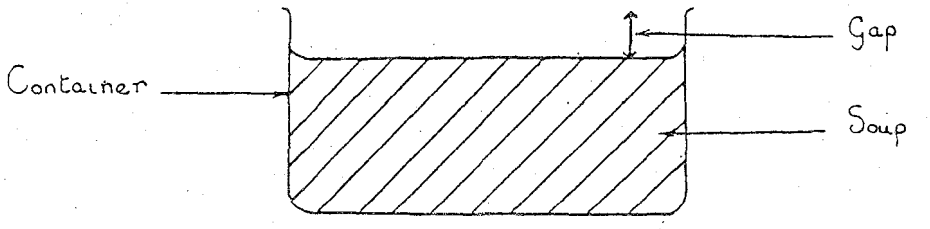
Rock T is called.....

Mark scheme

R—Haematite		1
S—Slate		1
T—Granite		1
	Total	③

(4) Question B-1496

When you freeze soups, sauces or stews, you are always advised to leave a gap at the top of the container.



a) Why is this gap necessary?

b) Why do vegetables and fruit only need a very small gap left in the container?

mark scheme

a) As water in the food freezes		1
it expands		1
b) Fruits have air spaces inside the cells		1
Fruits do not expand so much		1
Fruits have air spaces between them		1
	Maximum	⑤

(5) Question B-1566

People living in hot countries sometimes store their drinking water in earthenware pots. These pots are *porous*, which means that they have tiny holes which the water can pass through. Water stored in these pots keeps cooler than water in ordinary pots.

Say why a *porous* pot can help keep water cool.

.....

.....

mark scheme

Water fills the pores	1
A large surface area of water is therefore exposed to the atmosphere (surface of pot wet)	1
This increases the amount of evaporation which causes cooling (of the surface of the jar)	1
Maximum	③

(6) Question B-3491

You have been given two different minerals labelled P and Q. Look carefully at each mineral. Your job is to find a number of ways in which the two minerals are alike or are different.

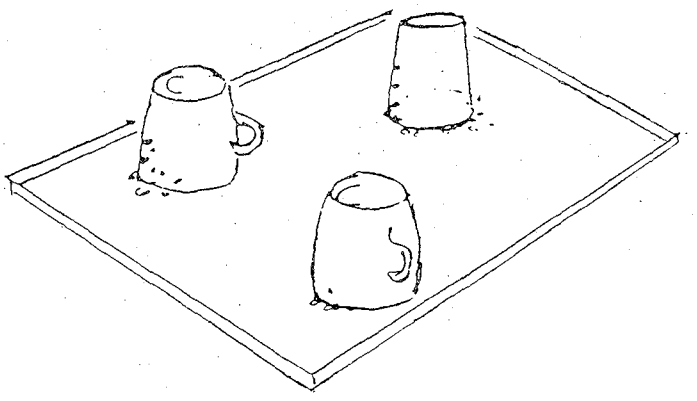
a) Write down *five* ways in which the minerals are different. Take each difference in turn and describe the difference for each mineral.

b) Write down *three* ways in which the minerals are alike.

mark scheme

a) Score up to 5 differences awarding		
4 marks for a complete and correct *comparison*	4	
1 mark per specific observation *not* made comparative	1	
		20
Record score for differences under two headings—		
A—score for comparative statements		
B—score for single specific statements		
b) 2 marks for each similarity. 3 similarities	2	6

(7) Question B-4071



Terry was washing up the dishes in hot soapy water.

He put the cups and glasses upside down on a tray to drain.

When he tried to lift them up to dry them he found they seemed to stick to the surface.

a) What happens to the air in the glass when the glass is first put down on the surface? (Remember that the washing-up water was *hot*).

b) Explain why it is hard to lift the glass up after a few minutes.

mark scheme

a) Trapped air expands	1
and some escapes	1
b) Air cools	1
contracts	1
excess air pressure holds glass down (or partial vacuum gives suction effect)	1
Maximum	⑤

(8) Question B-4077

When food is packed in bags for freezing it is important to remove all the air from the bag.

One way of removing the air is to plunge the bag up to the neck in a bowl of water before it is sealed up.

How does this remove the air?

.....

.....

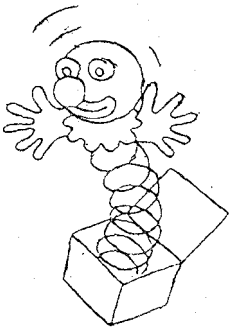
.....

mark scheme

Pressure under water is greater than air above	1
Extra pressure drives out air inside the bag	2
or Pressure of the water drives out air in the bag	2
Uses 'Force' rather than 'pressure' (e.g. force of water drives out air)	1
Maximum	③

(9) Question B-4163

When you undo the catch of a Jack-in-the-box, 'Jack' has enough energy to spring out.



Where is the energy before the box is opened, and how did it get there?

.....

.....

.....

mark scheme

(i) In the spring		1
(ii) Transferred from your muscles		1
(iii) —in compressing the spring		1
(in (iii) there must be the implication that *someone has compressed* the spring)		
Maximum		③

Code as follows: score only if *energy* is mentioned.

(i)	(ii)	(iii)	
1	1	1	A
1	1	0	B
1	0	0	C
0	0	0	D
0	0	1	E
0	1	1	F
1	0	1	G
0	1	0	H

(10) Question B-4168

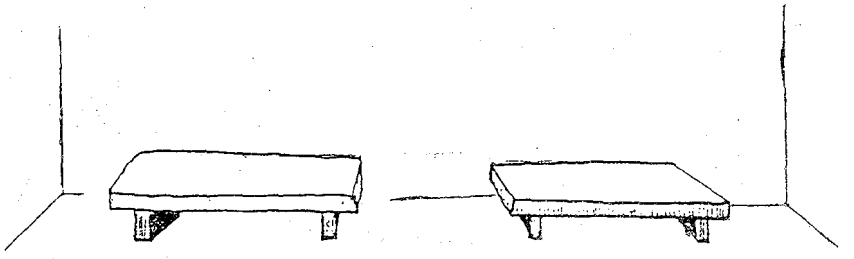
Some friends had found a very deep well in an old farmyard.
They were dropping pebbles down the well and listening for the splash.
They noticed that even small pebbles made guite a splash.
Tom said "It's funny that a small pebble has enough energy to make such a big splash".
Suggest one simple way in which Tom could have given the pebble even more energy, and say where the energy would have come from:

mark scheme

He could have			
(i) thrown not dropped		1	A
or			
dropped from a greater height		1	B
or			
thrown it up in the air first		1	C
If more than one of these		1	D
(ii) pebble has extra kinetic energy:			
energy from his muscles			
food			
plants			
sun			
(Any 2 or more from chain)		2	
	Maximum	③	

(11) Question B-4209

There were two benches at a station, exactly the same except that one was made of wood and the other was made of metal. In the winter, people soon felt cold when they sat on the metal bench; they didn't feel cold as quickly on the wooden bench.



Why did people feel colder sitting on the metal bench than on the wooden bench?

.....

.....

mark scheme

a) i) In winter both benches are at a *lower temperature* (colder) than people 1

ii) Metal is a *good conductor* of heat, wood is not 1

iii) So *heat is conducted away from* people who sit on the bench 1

Maximum ③

	1i)	ii)	iii)	
	1	1	1	A
	1	1	0	B
	1	0	0	C
	0	0	0	D
	0	0	1	E
	0	1	1	F
	1	0	1	G
	0	1	0	H
No response-				N