

TESTING FOR AUTHENTIC SCIENTIFIC UNDERSTANDING : AN INFORMATION PROCESSING APPROACH

YU — JEN SU

ABSTRACT

Science, by its very nature, is an information-intense activity having a mass of factual content, organizing concepts, specialist language, problem-solving skills, manual skills, observational skills and interpretive techniques. Even the simplest exam paper can contain components of all of these with the resultant risk of working memory overload and consequently it may no longer measure what we intend it does. It should never be taken lightly to set and score exam papers.

Based upon the currently popular perspective of information processing theory, there are three factors which control a pupil's ability to interpret and handle questions : (a) the limitation on the capacity of the working memory, (b) the noise which swamps the signal, and (c) the language or the words used in an item. This study is trying to interlink theoretical arguments with empirical evidence mostly conducted in the Centre for Science Education, University of Glasgow, Scotland. Starting from a brief overview of the information processing theory, this article explores subsequently the afore-mentioned factors which have effect upon the pupils' performances in the exam situations.

Research findings seem to be in favour of the arguments based on information processing model and they call for attention that we make sure to measure the authentic knowledge of science subject rather than a set of psychological factors. When setting a test question, the following guidelines are recommended :

1. What am I trying to test ?
2. How much information has to be manipulated ?
3. Does the student have to penetrate a noisy situation to get at what the question is asking ?
4. Is the noise justified by the objectives being tested ?
5. Am I aggravating the problems above by the words in the item ?

Introduction

The examination is a time-honoured feature of the teaching and learning world. In Taiwan, it has always been an extremely important part of academic selection and social placement, but seldom is the quality of university examinations explored in depth empirically to the extent that the problem-solving processes of students in an examining situation are uncovered systematically.

In any examining situations, information about subject matter is being processed by teachers as well as students. When setting an exam paper, the teacher has to present the material which he understands in a way in which students will also understand. The student then has to interpret the test items in terms of his personally constructed knowledge. When students come up with crazy ideas in the exam paper that we are sure we have never taught them, they have almost certainly been responding or reconstructing knowledge in line with their own rationalization. Much work has been done in recent years to understand the problem-solving processes of students in an examination. (Nussbaum and Novak, 1976; Johnstone and El-Banna, 1986)

The Glasgow school of information-processing has been devoted to the problem-solving processes in examinations for a long time and the working model we have advanced serves pretty well to account for the different performance of tests and the related cognitive factors in the thinking processes. This paper is trying to provide an overview of our research findings about examinations and to offer some useful messages for teachers and examiners in accordance with information processing theory.

A Working Model About Examining Processes

Figure 1 (Johnstone et al., 1982) shows a model of the reception, storing and retrieval processes in learning and we would do well to remind examiners of its salient features. In the middle of this diagram lies the short term/working memory which Miller (1956) reminds us is of very limited capacity. It is in this area of memory that human conscious processing undergoes by holding information and operating on it simultaneously. Incoming information from the senses has to interact with existing information recalled from the long term memory in human cognitive processes. The results of this interaction may be stored back in the long term memory in various forms or they may be exhibited in some sort of overt response such as writing, speaking and drawing.

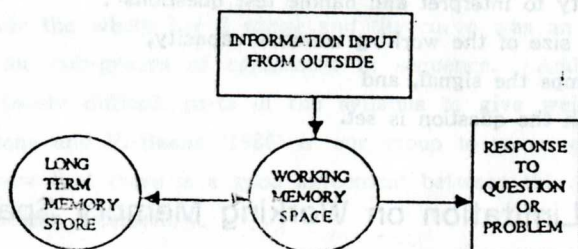


Figure 1: Information processing model.

If a candidate is asked a question in an examination, he reads and interprets that question into his working memory. He then has to add to this information from his long term memory store : facts, methods, strategies and so on. If this happens well within the space available in his working memory capacity, the pupil has the potential to respond correctly to the question. However, this is a necessary, but not sufficient criterion for successful performance because the pupil may recall or interpret wrongly or may not be inclined to bother to answer.

And what will happen if the working memory of the candidate is overloaded by the complexity of the test items ? Johnstone and Mughol (1979) argued that three things may result :

1. The pupil may find the question impossible and not answer in any more than a perfunctory fashion. This is the common response.
2. The pupil may find a way of breaking the problem down and dealing with a small portion at a time, each portion not exceeding the working memory space available to him.
3. The pupil may have developed a strategy which enables him to take short cuts by grouping information into readily processed parcels or "chunks".

The most important point of this model is that the material held in short term memory and the thinking activities have to share the same very limited space. To allow room for thinking, we have developed various strategies by which the recalled and new material is grouped (chunked) into associated units and each unit (or chunk) occupies a space in the scarce storage room of the working memory. These strategies are the product of our experience, our interests, our former learning or the stage of our development. If we do not have these chunking strategies, each piece of incoming and recalled information occupies a precious space in our working memory leaving little or no space for thought and so correct response is made difficult if not impossible. To make the most economical use of our memory and thinking space, a pupil can adopt the following techniques : either he limits himself to taking in only a small number of ideas at a time or he copes with the problem by calling on some prelearned strategy to organize the incoming information into manageable packages which are then handled in some sequence.

If we accept this model as a working framework, there are then some profound consequences for the way we make the setting and scoring of examination papers. There are several important cognitive factors about information processing in students' responses to test items, which we cannot ignore.

Based upon the information processing model illustrated above, there are altogether three factors which control a pupil's ability to interpret and handle test questions :

- (a) the limitation on the size of the working memory capacity,
- (b) the noise which swamps the signal, and
- (c) the language in which the question is set.

The Limitation on Working Memory Space

A few years ago our research group in the Chemistry Department of Glasgow University asked the Scottish Exam Board for the statistics for a set of their questions on a notorious portion of the syllabus of the dreaded MOLE (Johnstone and Mughol, 1979). We then analyzed each question for its demand in terms of what the pupil had to consider, recall and process before deciding how to solve this question. A plot of facility index against this demand was made and yielded the graph shown in Figure 2.

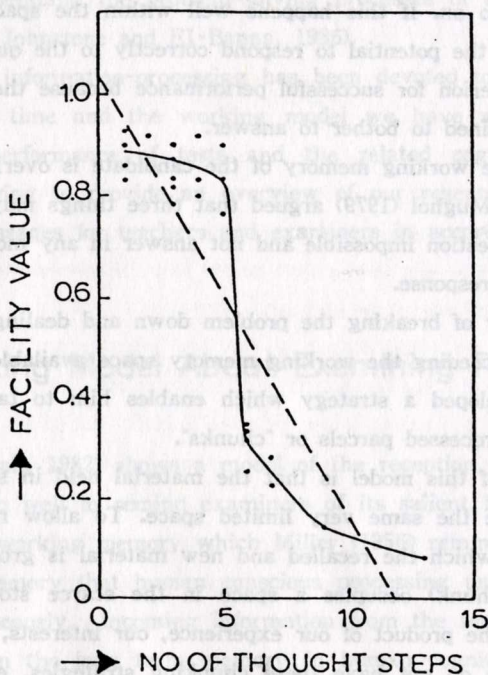


Figure 2: A plot of facility value in objective chemistry questions vs the number of thought steps needed to solve the question.

The attempt to find a line of "best fit" was not successful and a curve emerged, reminiscent of a pH curve. The interesting thing is that the vertical part of the curve came between five and six on the question complexity axis; tantalisingly close to 7 ± 2 . Clearly one group of questions was accessible to the pupils while another group was successfully done by only a few students. The sudden drop between these groups coincided with the afore-mentioned value which exactly gains its support from other psychological evidence that the working memory space is capable of processing about 5 ~ 7

pieces of information at a time.

The sample of some 20000 candidates would, of course, be made up of pupils whose working memory space spread over the whole 7 ± 2 range and the curve was an accumulation of a series of catastrophies occurring for sub-groups of candidates in sequence. And similar curves have been obtained for other notoriously difficult parts of the syllabus to give weight to our argument. It is made possible by Johnstone and El-Banna (1986) in our group to measure this mental working space for individuals and to show that there is a good agreement between this value and the complexity of the questions they can handle successfully.

If our model has any validity, it must contain a message for all who set questions. In assessment procedures the purposes of questions with high complexity will have to be reconsidered. If an exam paper contains questions with a range of demands, then not all questions will be accessible to all students. In other words, for most pupils such items simply measure the students' working memory capacity in the sense that we are no longer examining chemistry or science. Those of low working memory space will not be able to process the questions of high demand unless they have certain chunking strategies. This should be fairly reflected in overall scores.

Table I might help to illustrate this point further. It shows the results for pupils in an "O" Grade examination in terms of their measured working memory capacity. From this table it can be found that those of low capacity ($X=5$) have significantly fewer high grade passes and more failures (D and E) than those of high capacity ($X=7$). Are we then testing chemistry with our high demand questions or are we merely making psychological measurements of working memory? It should be noted even if an exam paper gives a reasonable spread of scores and a recognizable bell-distribution, there may be no true discrimination between mean scores for the three pupil groups (those of capacity 5,6 and 7) without using processing capacity as a discriminator.

GRADES GROUPS	A	B	C	D,E (fails)
X = 5	40%	20%	10%	30%
X = 6	58%	13%	10%	19%
X = 7	74%	14%	6%	6%

Table I: Percentage of students in each grade of the O-Grade chemistry examination (June 1985) corresponding to measured mental working space X.

The Irrelevant Noise

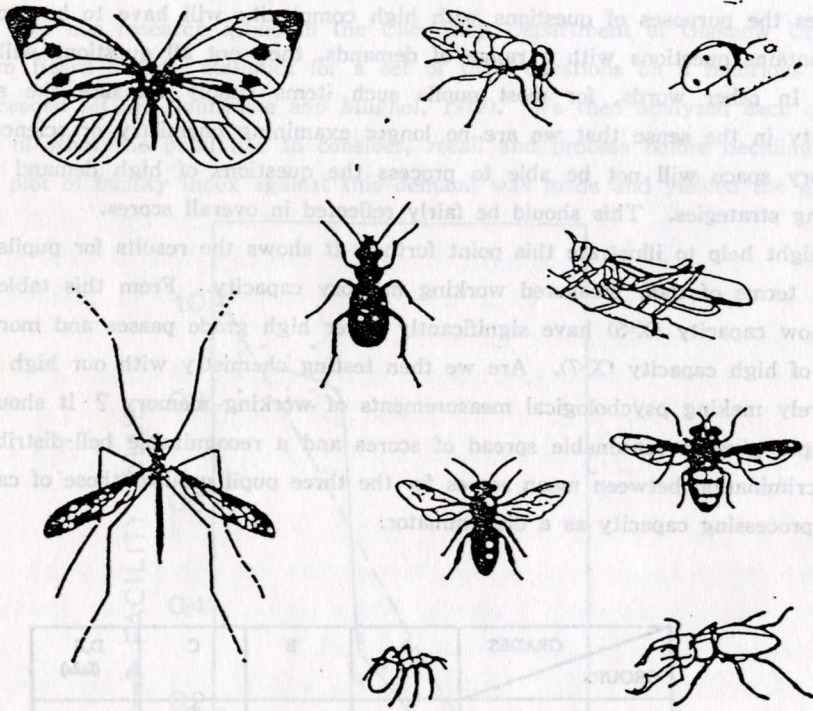
In most examining situations pupils are subjected to irrelevant "noise" which distracts them away from the "signal" of what is to be examined. Under normal classroom teaching, any teacher who is trying to communicate intelligently will not add "noise" consciously to obscure the "signal". However, when it comes to the examination, it is fatally easy to do so. Especially in most of Asian countries,

examinations are so popular that examiners generally add deliberate noise to obscure the signal.

According to the model described above, it should be noted that the distracting noise will never fail the pupils only because they are directed away from the signal. A lot of exam papers abound with such notorious "noise" and allow me to exemplify this by reference to a few test items (Question 1 and 2)

QUESTION 1.

There is a vast number of different varieties of insect on the Earth but they are all relatively small.



(a) Give a reason why very large insects do not occur.

 (2)

(b) Two similar groups of caterpillars are kept at different temperatures, one at 20°C and the other at 30°C.
 What differences could be observed between the two groups?

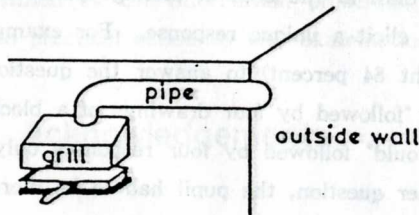
 (2)

In Question 1, the "noise" is anticipatory. This item is set to test data handling skills and students are expected to ignore all irrelevant data in order to look for patterns and relationships. However, what actually happened to the students sitting examination? A page full of pretty drawings of insects is getting the students ready to study these to look for possible response. But it turned out

that the question has no connection with the diagrams at all, leaving the students puzzled and worried. This kind of noise is quite different from that encountered in data handling questions, so great care has to be taken to ensure that students are not taken too far as to get lost in irrelevant noise jungle.

QUESTION 2

The owners of a cafe fitted a cooker hood over the grill, to carry away the smells and hot greasy vapour from the cooking food.



When they came to clean it out, they found that grease had condensed on the inside of the pipe.

A Why does the grease vapour condense on the inside of the pipe.

When the pipe was scraped clean, they found that the grease near the hood was sticky but the grease further along the pipe was runnier.

B Explain why the runnier grease has collected towards the end of the pipe further away from the grill.

C Describe two properties which a pipe would have to have if it was to be used for a cooker hood. Explain why each property is needed.

D Suggest one material which would be suitable for making a cooker hood.

Question 2 contains "noise" of a different kind. This question presumably is meant to be testing chemistry and it focuses on an apparently everyday situation. Students are expected to penetrate the situation to find out what is being asked. Chemistry knowledge shows that molecules of fat are broken into smaller molecules of assorted sizes during the heating process and then migrate along the pipe and condense at different temperatures, the bigger molecules condensing first. Big molecules tend to condense to give solids but small molecules tend to condense as liquids. But it is a situation which most students certainly will not have met unless they work in a restaurant and have the job of cleaning out the pipe. And what is more, in real life these small unsaturated molecules tend to join together to form hard waxy solids in the cold part of the tube. Once again, in this question the "signal" has been swamped by the "noise".

Language and Examination

Words in science teaching and examination are of obvious importance but Cassels and Johnstone (

1983) found that the words tested in multiple choice test items of various kinds with distractors required a low level of discrimination. Language itself has many problems associated with it which can add significantly to the load that has to be borne by the working memory.

In a study supported by the Royal Society of Chemistry in the Great Britain (Cassels and Johnstone, 1980), it was shown that the non-technical words associated with science were a cause of misunderstanding for pupils. Words which were understandable in normal daily life changed their meaning when transferred to a science situation. Even teachers in training found difficulty themselves with words like average (which they took to mean "ordinary"), converse, criteria, theory and valid.

Words themselves do not have an intrinsic difficulty level, but depend very much upon context. This is nowhere more evident than in the area of testing, particularly objective testing, where precision of language is essential to elicit a unique response. For example, pupils found it more difficult (39 per cent correct compared with 84 percent) to answer the question "In which of these diagrams is the block of wood immersed?" followed by four drawings of a block of wood and water, rather than "If we immerse an object we would" followed by four responses only one of which involved a liquid. To be able to answer the former question, the pupil had to be more precise in his understanding of "immerse".

Questions set in a positive form brought better performances from pupils than negative ones. If questions contained double negatives (one in the stem and one in the options), the performance was very poor. Wason (1959,1961) has shown that ideas in a negative form occupy twice as much space in the working memory as positive forms. Double negatives may even occupy four times the space occupied by a positive form. It is little wonder that negative questions fare so badly in tests in that they leave less space in the working memory for thought.

The information processing model strongly supports our arguments because the differences of performance in the questions could be substantially accounted for by the thinking processes necessary to solve the question.

Conclusion and Suggestions

It is very hard for a teacher to recall what it was like to sit an examination for the first time. We have gained expertise and confidence constantly which makes us see as trivial, things which are far from trivial for a pupil answering an exam paper. It is important to realize that the pupil finds himself in constantly grappling with our question in an examination. The "clever" question may be so clever that it no longer measures what we think it does. It may be in conflict with a set of psychological factors and it is these that we are measuring rather than the subject matter.

When we are setting an exam paper, we could ask some very clear check questions about each test item. In summary, based upon most of the discussion in this paper about examining, I would recommend all teachers at all levels to :

- (1) define clearly what I am trying to test;
- (2) check each item thoroughly how much information has to be manipulated;
- (3) ensure that the student will not be distracted by a too noisy situation to get at what the ques-

tion is asking;

- (4) consider if the noise is justified by the objectives being tested;
- (5) make sure that you are testing what you have intended to without undue testing of other factors;
- (6) check if you are aggravating the problems by the words in which the test item is set.

In an examination situation except an oral exam, the pupil has no opportunity to interchange their understanding or opinions. He has to interpret the answer himself without the benefit of a two-way exchange of ideas. This makes the setting and scoring of exam papers an extremely hazardous job which should by no means be taken lightly. The ideas mentioned in this paper have been empirically evidenced and theoretically justified by the information processing theory. We believe, if well adopted and applied, they can lead to practical action by any examination setters.

Acknowledgements

The Author is much obliged to my former supervisor Professor A.H. Johnstone, Director of Centre for Science Education in the University of Glasgow, U.K. Without his inspiring edification and generous permission, the completion of this article would not have been possible.

References

- Cassels, J.R.T. and Johnstone, A.H. (1977). "Language in chemistry", Report of Education Division Conference, Research for the classroom and beyond, 48-54. London: The Royal Society of Chemistry.
- Cassels, J.R.T. and Johnstone, A.H. (1980). Understanding of Non-technical Words in Science. London: the Royal Society of Chemistry.
- Cassels, J.R.T. and Johnstone, A.H. (1983). The meaning of words and the teaching of chemistry. *Education in Chemistry*, 20(1), 10-11.
- Johnstone, A.H. (1984). New stars for the teacher to steer by? *Journal of Chemical Education*, 61(10), 847-849.
- Johnstone, A.H. and El-Banna, H. (1986). Capacities, demands and processes - a predictive model for science education. *Education in Chemistry*, 23(3), 80-84.
- Johnstone, A.H. and Mughol, A.R.N.(1979). Testing for understanding. *School Science Review*, 61(214), 147-150.
- Johnstone, A.H. and Wham, A.J.B. (1979). A model for undergraduate practical work. *Education in Chemistry*, 16, 16-17.
- Johnstone, A.H. and Wham, A.J.B. (1982). The demands of practical work. *Education in Chemistry*, 19(3), 71-73.
- Miller, G.A. (1956). The magic number seven, plus or minus two. *Psychological Review*, 63, 81-97.
- Nussbaum, J. and Novak, J. (1976). An assessment of children's concepts of the earth utilizing struc-

tured interviews. Science Education, 60, 535-550.

Su, YU-JEN. (198). How to improve problem-solving ability in chemistry. Science Education Bimonthly, (in chinese).

Wason, P.C. (1959). The processing of positive and negative information. Quarterly Journal of Experimental Psychology, 11, 92-107.

Wason, P.C. (1961). Response to affirmative and negative binary statements. British Journal of Psychology, 52, 133-142.

Acknowledgements

The Author is much obliged to my former supervisor Professor A.H. Johnston, Director of Centre for Science Education in the University of Glasgow, U.K. Without his inspiring education and generous permission, the completion of this article would not have been possible. Thanks are also due to my colleagues at the University of Glasgow for their support and interest in this project.

References

Cassels, J.R.T. and Johnston, A.H. (1987). Language in chemistry: Report of Education Division. Conference Research for the classroom and beyond, 48-54. London: The Royal Society of Chemistry.

Cassels, J.R.T. and Johnston, A.H. (1980). Understanding of Non-technical Words in Science. London: The Royal Society of Chemistry.

Cassels, J.R.T. and Johnston, A.H. (1983). The meaning of words and the teaching of chemistry. Education in Chemistry, 20(1), 10-11.

Johnston, A.H. (1981). New stars for the teacher in error. 84-91. Journal of Chemical Education, 58(10), 681-682.

Johnston, A.H. and Johnston, J.E. (1986). Capabilities, demands and processes in problem-solving for science education. Education in Chemistry, 23(3), 80-84.

Johnston, A.H. and Wharm, A.L.B. (1982). A model for understanding practical work. Education in Chemistry, 19(1), 16-17.

Johnston, A.H. and Wharm, A.L.B. (1982). The demands of practical work. Education in Chemistry, 19(3), 71-73.

Miller, G.A. (1956). The magic number seven, plus or minus two. Psychological Review, 63, 81-97.

自然學科測驗的新視窗 (訊息處理論對測驗的貢獻)

蘇 育 任

摘 要

自然科學本質上說來，是一種訊息強度很大的活動，涵蓋了科學知識、組織的概念、專屬術語、解決問題的技能、操作、觀察與解釋等技能。再簡單的一份自然科試卷，都可能包含上述各種成分，以至於測驗時極易使學生的腦筋負荷過重，而無法真正測量學生對科學知識的理解，故自然科測驗的命題與評分必須十分小心。

根據目前頗為風行的訊息處理論，考試時學生解答考題的能力乃是受三個認知因素影響，即：①個體工作記憶容量的大小，②考題中不相干的雜訊掩蓋過主要訊息，和③題目中所使用的術語及文字敘述。本研究旨在介紹英國格拉斯哥大學科教中心，所進行的一系列科學測驗與評量之成果，首先簡要地探討訊息處理論，接著依據這理論，逐一探討學生在考試時，受那些認知心理因素的影響。

概括說來，實徵性研究結果與訊息處理論的推測十分吻合，他們呼籲測驗學者或教師們要注意設計科學測驗時，必須遵照下列建議，以免試題設計不當，變成只測學生的心理能力而不是在測其學科成就。茲將命題時的一些注意事項臚列如下，以供參考：

1. 此題究竟要測驗什麼？
2. 要解這道題目，學生須操作的「訊息數量」有多少？
3. 學生是否會被本題目中太多「無關的雜訊」迷惑，而找不到真正的答案？
4. 題目中放入的誘答，是否合理？
5. 題目的文字敘述、專有名詞等是否合宜？