CHEMISTRY EDUCATION: RESEARCH AND PRACTICE IN EUROPE
2001, Vol. 2, No. 3, pp. 313-327

THE PRACTICE OF CHEMISTRY EDUCATION (PAPER)
Assessment

Alex H. JOHNSTONE and Abdullah AMBUSAIDI
University of Glasgow, Centre for Science Education

FIXED-RESPONSE QUESTIONS WITH A DIFFERENCE

Received 14 August 2001; accepted 15 October 2001

ABSTRACT: As we have thought about objective, fixed-response questions there have been recurrent themes of unease. In a previous paper in *this journal* (*CERAPIE* 2000,Vol. 1, No. 3, pp. 323-328) we set out evidence from the literature and from our own research, that the “conventional” forms of fixed response questions had serious drawbacks. This paper offers three other types of fixed-response questions which are designed to overcome these problems, at least in part. [*Chem. Educ. Res. Pract. Eur.*: 2001, *2*, 313-328]

KEY WORDS: *assessment; objective testing; higher education.*

INTRODUCTION

When a student makes a choice in a fixed-response situation, is the correct choice being made for a right or for a wrong reason? In our previous paper we quoted the findings of Tamir (1990) which showed that about one third of students choosing the correct option in a multiple choice question did so for a wrong reason. Even when a wrong choice is made, is it out of ignorance or for a very good reason? None of the “conventional” fixed-response types, such as multiple choice, allows the examiner to ascertain this reasoning background. Assumptions are made that if a student makes a wrong choice, this necessarily indicates a certain lack of knowledge or betrays some specific confusion. However, if the differences between the best answer and the distracters are subtle (as they should be at tertiary level), there is some degree of “rightness” in all the distracters and students who have read more, or thought more, might choose a distracter instead of the “correct” answer for very good reasons. Unfortunately, the nature of these questions does not provide a student with the opportunity to “tell all”.

A second part of our unease stems from the fact that students who have moved on to levels of intellectual maturity, described as Perry C (Johnstone, 2001), find fixed-response questions irksome, confining and frustrating and they often choose distracters rather than the “correct” answer for very good reasons. This sometimes shows as *negative discrimination* for some questions in which the success rate for the top third of the class is *less* than the success rate of the bottom third! The more discerning students have chosen distracters other than the “correct” one while the less insightful (or more naive) score well because they cannot see the other possibilities.

Lessening of the disadvantages of fixed-response questions

All of this presents real problems for the use of fixed-response questions unless other types can be found which provide the advantages of fixed-response (e.g. rapid and reliable marking) with a lessening of the disadvantages.

What about recognition for partial knowledge?

Our research (Friel & Johnstone, 1978) has shown that if the same area of learning is assessed by normal open-ended, methods and also assessed by objective, fixed-response methods, two orders of merit are generated for a given group of students. One might expect that, since the same knowledge and understanding is being assessed, the two orders of merit should be substantially the same for the same sample of students. The best student by one method should be the best by another method and so on down the line. But experimentally this *is not so.* If a correlation is worked out between the two orders of merit, it usually comes out at about 0.6. This figure turns up frequently in the research literature. For those not familiar with a numerical value for rank-order correlation, a word of explanation may be necessary. A perfect match in order would result in a value of 1.0; a complete reversal of the order would give a value of -1.0. A completely random pair of orders would give a value of zero. The experimental value of 0.6 suggests that the two orders of merit have some similarity, but are by no means well matched.

This drives us to ask why the orders do not match. Much experimental work has been done to try to answer this question and the most important factor to emerge is that, in scoring the open-ended questions, credit is given for partial knowledge or for wrong conclusions arrived at for good reasons. In the fixed-response situation, no such credit is given. This brings us back to the problems we raised at the beginning of this paper about the lack of evidence for student reasoning.

There have been several ingenious attempts made to score multiple-choice questions to allow for partial knowledge. Some of these ask the students to rank all the responses in the question from the best to the worst. In other cases students are given a tick (√) and two crosses (×) and asked to use the crosses to label distracters they know to be wrong and the tick to choose what they think is the best answer. They get credit for eliminating the wrong, as well as for choosing the correct. These are obviously more difficult to score, but worth the effort. The rank order produced when these devices are applied to multiple-choice tests and the rank order produced by an open-ended test correlate to give a value of about 0.9; almost a perfect match. This underlines the importance of the examiner having the means of detecting and rewarding reasoning.

What about guessing?

Guessing has always been a problem with fixed-response questions and various strategies have been proposed for reducing or even eliminating the effects of guessing. However, they are outside the scope of this paper, and for this reason not discussed here.

What about other forms of fixed-response questions?

Thinking about forms of fixed-response questions has turned to:

- giving credit for partial knowledge;

- reducing the possibility of guessing;
- finding indications of reasoning paths.

The remainder of this paper will be devoted to examining three fixed-response formats which attempt to make allowances for the weaknesses of the conventional formats. These methods are not new, but have been largely neglected. We claim no originality for them, but we believe that we have been able to develop them into forms which have been found to be useful in the assessment of chemistry at all levels and so we should like to share them with colleagues who are looking for new assessment tools.

OTHER FORMS OF FIXED-RESPONSE QUESTIONS

Interlinked True/False Questions

Conventional True/False questions, where each true/false decision stands alone and is independent of the questions before or after, are well known. They are widely used by Medical Schools for assessing students. They are open to criticism on the grounds of blind guessing and for giving no indication of reasoning. Elaborate scoring systems have been devised to discourage guessing, but they are very suspect. Students have been heard to refer to them as "multiple guess"!

In an interlinked format, each true/false decision has consequences for the next decision and so on along a chain. Let us take a stylised example (Figure 1).

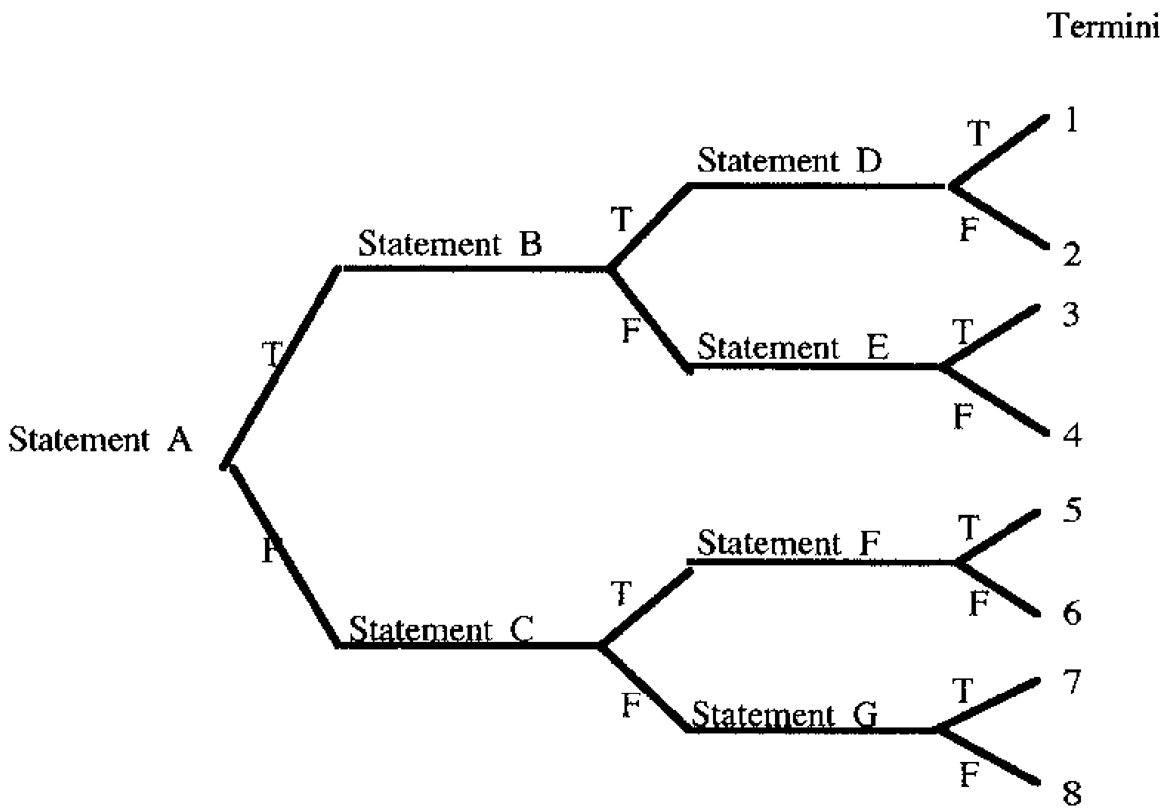

FIGURE 1. *Interlinked true/false format.*

This is best done on a computer, but paper methods are possible by which each decision directs the student to another page where the next question appears.

Every student is presented with statement A and asked to pronounce it to be either true (T) or false (F). If the student decides that it is *true*, he is presented with statement B which is in some way a consequence of this decision. If, however, the student decides that statement A is *false*, he is directed to statement C which is, in some way, a consequence of this decision. This process continues from statement B to either D or E or continues from statement C to either statements F or G. Finally a decision is made about D (or E) or about F (or G). Each student makes decisions on only three statements, each one a consequence of previous decisions. This brings the student to a terminus (numbered 1- 8), each one of which *uniquely defines the route taken* by the student. For example, a student arriving at terminus 3 has declared statement A to be true, statement B to be false and statement E to be true.

Let us suppose that the "best" answer is to arrive at terminus 4. (Students arriving here purely by guesswork would have a one in eight chance of doing so.) Terminus 4 would get the best score, terminus 3 would get credit for two correct decisions on the way and each of the other termini could be given partial scores. Even students who arrived at termini 5 - 8 would get credit for correct decisions on the way, although their initial decision on statement A was wrong. Also, each terminus could carry diagnostic and remedial help for each student.

An example of a set of linked true/false statements is shown below (Figure 2).

Consider the following three statements: (1) Oxidation means loss of electrons. (2) In ALL oxidations the positive charge on the ion increases. (3) $Cr_2O_7^{2-} \rightarrow CrO_4^{2-}$ *is an oxidation. For each of these statements, decide if it is true or false.*

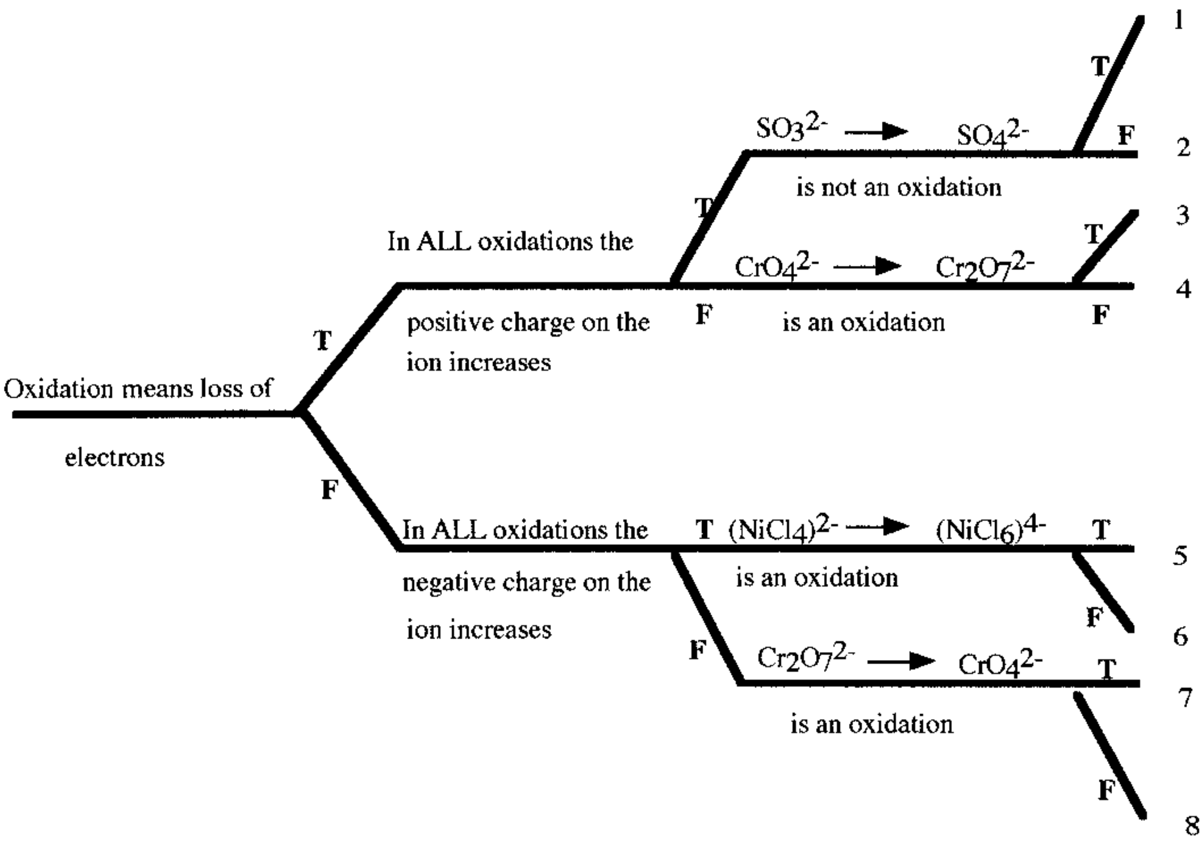

FIGURE 2. *A set of linked true/false statements.*

This is attempting to explore misapprehensions which are held about Oxidation and Reduction. The "best answer" would be *true* (*oxidation means a loss of electrons*) followed by *false (in all oxidations the positive charge on the ion increases)* and finally *false (CrO_4^{2-} giving $Cr_2O_7^{2-}$ is an oxidation),* but other combinations of "rightness" would merit some credit. This kind of question set is best administered by computer and then each choice-point (*or node*) can be made even more sophisticated.

On the way through a true/false sequence, a student may realise that a wrong choice has been made further back and the possibility of back-tracking is needed. This is easily achieved by computer and the machine can keep a record of the student's choices and the exact route taken. This would give the teacher information by which the student can be helped later. Programs are available into which a teacher can fit a set of statements with no programing skill and so construct a range of interlinked true/false statements.

A further use of such interlinked true/false questioning is to prepare students for a discussion in which to explore a series of decisions and follow their consequences. For example, students can be presented with an industrial situation which may take several sentences to describe and are then asked to make a decision (sometimes against the clock). Depending upon the decision, a consequence appears to which they have to react with another decision. This can lead to yet another consequence and so on. A set of wrong decisions might lead to the destruction of the factory or a combination of rights and wrongs may require some drastic action to restore the situation and prevent disaster. Situations like this can be handled in groups to stimulate discussion or individually as a form of assessment.

Venn Diagrams

This is a simple, pictorial form of assessment which allows for degrees of "correctness" and is best used in situations which require the *ability to categorise*. These diagrams are used in teaching mathematics and other subjects to encourage a logical approach to categories, sub-categories and shared categories. Some examples are given below (Figures 3, 4, 5).

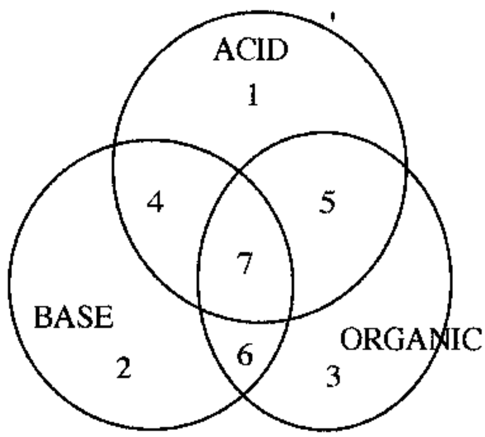

FIGURE 3. *One example of a Venn Diagram.*

Into which area of this diagram do the following chemical species best fit? Indicate your choice by writing the area number against the species name.

	Species	*Area number*
(a)	H_2O	
(b)	$C_2H_5NH_2$	
(c)	$(CH_3)_2O$	
(d)	$C_2H_5CO_2^-$	
(e)	$^-O_2CCH_2NH_2^+$	
(f)	NH_4Cl in liquid NH_3	
(g)	$BrF_2\ SbF_6$ in BrF_3	

The best answer for water would be area 4, because it can exhibit both acidic and basic properties as a proton donor or acceptor. However, a student may know that water can accept a proton to give H_3O^+ but may not know that it can release H^+ at a cathode. This student would choose area 2. The nature of the partial knowledge becomes obvious to the examiner. Similar evidence of partial or even wrong knowledge is made evident by the student's choice of area for each of the species listed. Scoring can be weighted to take account of these choices.

Another example might be:

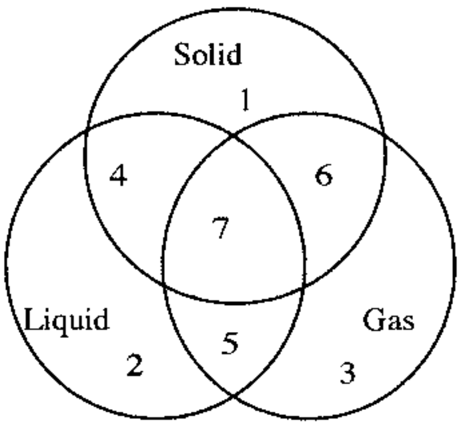

FIGURE 4. *Another example of a Venn Diagram.*

Substances at 0^o C and 1 atmosphere pressure may exist in one or more than one physical state: solid, liquid or gas. Indicate where each species would best fit in the Venn diagram by placing the area number against the species.

	Species	*Area number*
(a)	Water	
(b)	Iodine	
(c)	Benzene	
(d)	Naphthalene	

The Venn diagrams need not be a set of three intersecting circles. An example might be as in Figure 5.

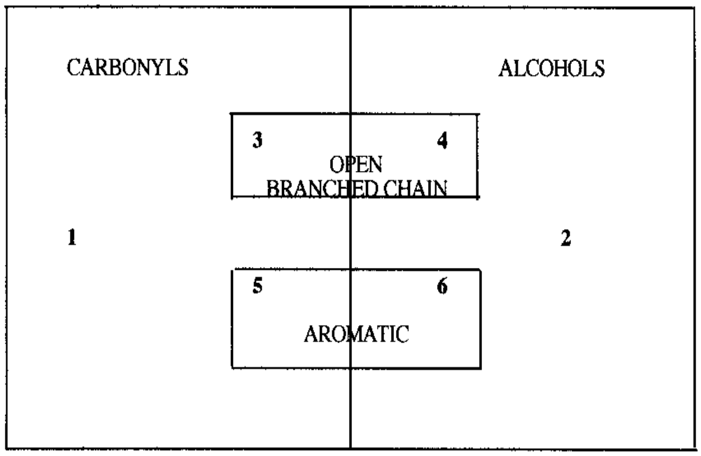

FIGURE 5. *A Venn Diagram not in the form of intersecting circles.*

Into which area in the diagram do the following compounds best fit. Indicate your choice by writing the area number against the name of each compound.

	Compound	*Area number*
(a)	1,2 dihydroxyethane	
(b)	1 phenylpropan-3-one	
(c)	2 hydroxy 3 methylbutane	
(d)	1,3 dihydroxycyclohexane	
(e)	cyclopentan-ol	
(f)	hexan 1,6 dial	

These questions are easy to set and to mark and they give an indication of partial knowledge. The large number of areas reduces the possibility of guessing.

Structural Communication

This is a very powerful and flexible method of fixed-response assessment which can range in use from the checking of facts and simple relationships to the construction of "objectively markable essays". The earliest ideas for this kind of assessment are found in the work of Egan (1972) and have since been developed and expanded by other workers, particularly in our research group. The name, Structural Communication, which Egan used, aptly describes how the method works. The students are presented with a random array of information and are asked to rearrange the array in such a way as to communicate their knowledge and understanding to the examiner. The structure, which the students impose on the information given, reveals their ideas and their reasoning. The array of information can be presented as a grid of numbered boxes each containing a piece of information, or the information may appear as a series of numbered statements presented one below the other. By a series of examples, the range of possible uses of the technique will be illustrated.

(a) *Categorising* (Figure 6)

1 IRON	2 ALUMINIUM	3 MAGNESIUM
4 LITHIUM	5 COPPER	6 LEAD
7 ZINC	8 BARIUM	9 TIN

FIGURE 6. *Structural Communication: Categorising.*

The grid contains the names of metals. Answer the following questions by selecting appropriate boxes and writing their numbers in the spaces provided. **Note** that any box may be used several times to answer different questions.

Question 1. Which of these are Transition Metals?
Question 2. Which of these alloy together to give Brass?
Question 3. Which of these react with cold water to give hydrogen?
Question 4. Which of these are Alkaline Earths?
Question 5. Which of these combined with oxygen, can form negative ions ?
Question 6. Which of these would be known in Roman times?

This is testing knowledge and relationships and is covering a wide scope.

There is no indication of only one correct response (as in multiple-choice), because one or more box numbers may be required to answer each question and so guessing is much reduced. The same box can be used several times as part of the answer to a number of questions and so answering by elimination is avoided. Partial knowledge is allowed for. However, there is one drawback which must be countered. If students are given credit for their correct choices and go unpenalised for wrong choices, they could give all nine boxes as the answer to all the questions.

Egan suggested a correction factor to get round this. Suppose that the correct answer to Question 1 above was IRON, COPPER and ZINC (i.e. boxes 1,5,7). There are three "correct" boxes and six "incorrect" boxes. His scoring system was:

$$\text{Score} = \frac{\text{Number of correct boxes chosen}}{\text{Number of correct boxes available}} - \frac{\text{Number of incorrect boxes chosen}}{\text{Number of incorrect boxes available}}$$

A student who responded to Q.1 above with 1, 5, 7 would have: Score = $\frac{3}{3} - \frac{0}{6} = 1.0$.

A student who responded 1 and 5 and omitted 7 would have: Score = $\frac{2}{3} - \frac{0}{6} = 0.7$ (Partial knowledge is rewarded). However, if a student's response was 1, 2, 7 the score would be given by: Score = $\frac{2}{3} - \frac{1}{6} = 0.5$ The student who chose all the boxes would have a score of $\frac{3}{3} - \frac{6}{6} = 0$.

This arithmetical procedure is a little tedious to do by hand, but computers mark this effortlessly. Students, and teachers, are not too happy with fractional scores, and so all of these can be multiplied by some simple factor such as ten to give whole numbers. And so the first student above would score 10 x 1 = 10; the second would get 7; the third get 5 and the last still get 0.

Programs are commercially available which remove any chore and apply any weighting the examiner desires (e.g. TRIAD, MacKenzie, 1997). A full treatment for scoring Structural Communication questions is given in the Appendix.

(b) Pattern seeking

This is an extension of categorisation in which the examiner gives examples and non-examples of some pattern and asks the student to deduce the pattern and seek for other examples (Figure 7).

1 $Fe^{2+} \rightarrow Fe^{3+}$	2 $Cr_2O_7^{2-} \rightarrow Cr^{3+}$	3 $CrO_4^{2-} \rightarrow Cr_2O_7^{2-}$
4 $Fe(CN)_6^{4-} \rightarrow Fe(CN)_6^{3-}$	5 $Al^{3+} \rightarrow AlO_3^{3-}$	6 $Cu^{2+} \rightarrow Cu_2O$
7 $S_2O_3^{2-} \rightarrow S_4O_6^{2-}$	8 $MnO_4^{-} \rightarrow Mn^{2+}$	9 $[Ni(H_2O)_6]^{2+} \rightarrow NiCl_6^{4-}$

FIGURE 7. *Structural Communication: Pattern seeking.*

In these questions about reactions you will be given two examples and one non-example of an idea the examiner has in mind. You are asked to work out this idea and show that you have found it by choosing further examples from the grid.

Question 1. Examples are in boxes 2 and 6, but box 5 is a non-example. Choose any other examples from the grid and write the number(s) here.

(The reasoning should be that 2 and 6 are examples of reduction while 5 is neither oxidation nor reduction. The student response should be box 8 because that is the only other example of reduction.)

Question 2. Boxes 3 and 5 are examples; box 4 is a non-example. Choose any other examples from the grid and write their number(s) here.

(The reasoning might be that 3 and 5 are neither oxidation nor reduction while 4 is an oxidation. The student response should be box 9, because that is the only example of neither oxidation nor reduction.)

(c) Sequencing

In this version, the student is asked to choose relevant boxes and then to sequence the responses to communicate more fully (Figure 8).

1 TIN	2 IRON	3 COPPER
4 BRASS	5 MAGNESIUM	6 LEAD
7 COBALT	8 BRONZE	9 SODIUM

FIGURE 8. *Structural Communication: Sequencing.*

Question 1 Select the metals from the grid which are TRANSITION metals and arrange them in order of *increasing* atomic number.

Question 2. Which metals are in the alloy in box 8? List them with the one of highest proportion first.

Question 3. Arrange the elements in boxes 2, 3 and 9 in order of their date of discovery (oldest first).

The scoring for this type of question is more complex because it must have two parts, the *choice* of boxes and the *sequence* order. The choice can be scored as explained before, but the sequence presents a problem. There are a number of methods available for computer marking. If the correct choices have been made (no more and no fewer), the computer compares the sequence with one provided by the examiner. A perfect fit gets full marks, a complete reversal gets no marks. If two adjacent responses are in the correct order, but the remainder are out of order, a partial mark is given. However, if the original choices were incorrect (for example, by the inclusion of a wrong choice) this is penalised by the first stage of the scoring and then ignored in the sequencing. This is tedious to do by hand, but is easily achieved by machine marking.

So far most of the examples have been presented as one word in each box, but the boxes can contain diagrams, pictures, sentences, formulae (mathematical or chemical) or structures. This increases their flexibility and usefulness (Figure 9).

1 CH_3OH	2 OCl | $CH_3\,C{=}O$	3 H | $CH_3C{=}O$
4 C_2H_5Cl	5 Cl | $CH_3CHC{=}O$ | OH	6 C_2H_5OH
7 $CH_3C{=}O$ | OH	8 $C_2H_5C{=}O$ | H	9 OCH_3 | $CH_3C{=}O$

FIGURE 9. *Structural Communication: Chemical formulae and structures.*

Question 1. Select the box(es) which contain alcohols.
Question 2. When the substance in box 6 is oxidised it can produce more than one product. Which boxes contain these products?
Question 3. If this oxidation is carried out in stages, arrange your choice of products in order of their occurrence.
Question 4. Select the box(es) which contain acid chlorides.
Question 5. Which boxes contain the substances which are used to make the compound in box 9?

Another example is shown in Figure 10.

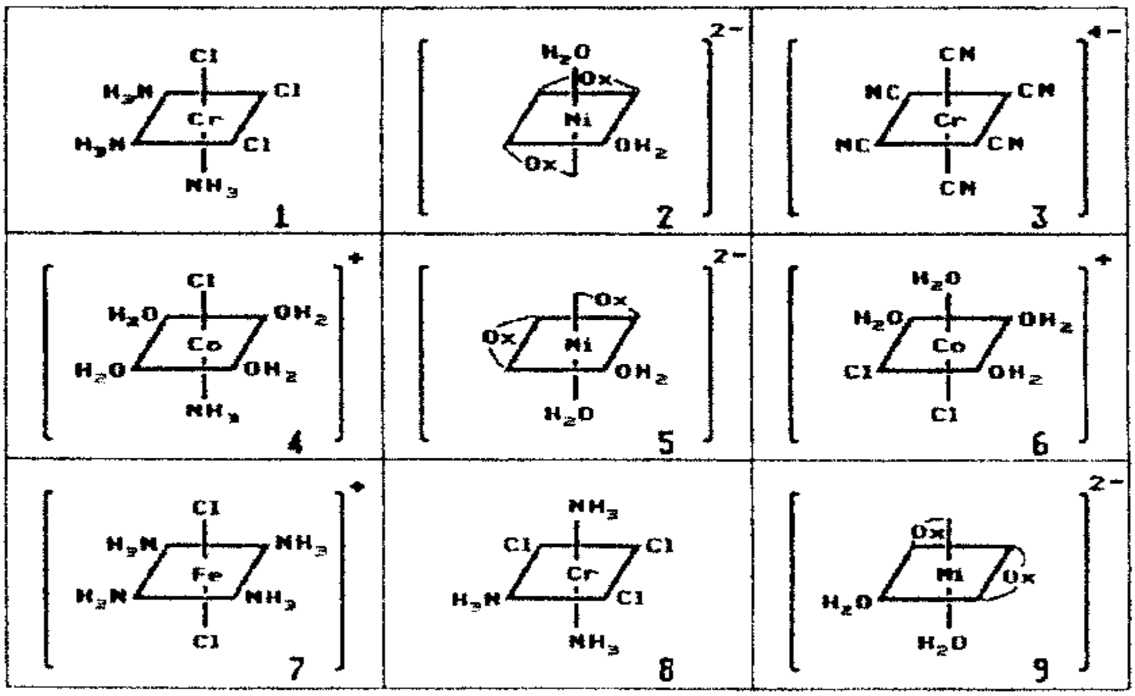

FIGURE 10. *Structural Communication: Another example with boxes containing chemical structures.*

In this grid the symbol "Ox" represents the oxalate (ethanedioate) ion which is a bidentate ligand.

Question 1. Which box contains the *fac*-triammine trichloro chromium(III) species?
Question 2. Which box(es) contain a d^3 species?
Question 3. Which boxes contain complexes which are chiral?
Question 4. Which two boxes contain a chiral pair?

(d) The "Objective Essay"

Conventional extended-answer questions are good for assessing reasoning and the ability to marshal material into a logical whole. They are, however, difficult to mark consistently. Something approximating to an extended-answer question can be achieved by Structural Communication techniques. Indeed Structural Communication is at its best when used in this way. The setting of such questions is most easily done in this way.

* Ask yourself a question which would need three or four sentences (or ideas) strung together to answer it. Reduce your answer to these basic ideas and insert them randomly into the blank grid.

* Now ask yourself a second and related question and proceed as before. Preferably some of the ideas needed to answer this second question were also necessary to answer the first.

* Finally, if you wish, ask yourself a third question related to the first two and complete the grid.

An example might look like this (Figure 11). You may need to extend the grid beyond the nine boxes we have used in the previous examples, but twelve, or at most sixteen, boxes can be used at university level.

1. If the ligands are weak (eg water), Δ_0 is small and electrons can move to higher energy orbitals.	2. Different ligands can change the colour of the complexes of a metal ion.	3. The difference in energy between the orbitals (Δ_0) corresponds to a frequency in the visible range.
4. Paramagnetism is caused by the presence of unpaired electrons.	5. When a Transition Metal ion is surrounded tetrahedrally by ligands, two of the orbitals have a higher energy than the other three.	6. Strong ligands like CN^- create a large Δ_0 and discourage electrons from moving to higher energy orbitals.
7. Transition Metal ions have *d* electrons which occupy orbitals of the same energy when the ion is in the free state.	8. If the ion has six *d* electrons, they can be arranged either as three pairs or as one pair and four singles.	9. White light shining on the complex can promote electrons by supplying energy equal to Δ_0. This energy is subtracted from the white light and we see the colour which is left

FIGURE 11. *Structural Communication: Assessing reasoning and ability to place material in a logical order.*

Using the pieces of information in the grid, construct answers to the following questions by choosing the numbers of the relevant pieces and presenting them in a logical order. This logical order should read as if you were writing a short essay.

Question 1. Explain why $Ti\,(H_2O)_6^{2+}$ is purple.
Question 2. Explain why $Fe(H_2O)_6^{2+}$ is paramagnetic while $Fe(CN)_6^{4-}$ is not.

At this level of sophistication, there may be more than one logical order permissible, but the computer (or even the hand marker) can cope with this. The scoring of the selection is easily done as before, but the scoring of the sequencing is more complex, but not too difficult. The possibility of guessing can be largely ignored. This is nearly a machine-markable essay, testing most of the skills which a normal essay requires; selecting relevant material from irrelevant and presentation in a logical order. Students do not find this an easy option to the conventional essay, but the marker's burden is considerably eased!

A form of Structural Communication, which is now available commercially (TRIAD, MacKenzie, 1997), presents the fragments as a succession of statements down the screen, with the question at the top. The student is asked to study each fragment and decide whether or not it is needed to answer the question. The student selects the fragments and all the others disappear from the screen for the time being. The full screen can be recalled if there is any doubt. The student then "drags" the fragments around the screen to obtain the logical sequence required and, when satisfied, declares the question answered. This is then repeated for each of the questions. Optically this is probably better than the grid, in that the student's final decisions are made without the intrusion of the irrelevant pieces.

CONCLUDING REMARKS

We have now looked at three techniques which try to avoid the main drawbacks of fixed-response testing: (i) guessing; (ii) lack of information about reasoning; and (iii) no allowance for partial knowledge. There is, however, one problem which has not been overcome and that is the need for freedom of expression by students who are at the more developed end of the Perry scale. They want to be able to show their own ideas and their own reasoning and have room for original insights. The remedy which makes assessment open and "congenial" to all students, is to use a mixture of assessment tools. Both open-ended and fixed-response testing have a place with the proportions changing in favour of the open-ended as the students mature. In fact, since students are driven by assessment, this change in the blend towards more open-ended testing may be one of the tools necessary to encourage the maturation.

Much effort is being expended on curricular innovation without the same effort being applied to assessment innovation. It is worrying to see academics, being driven by various forces to "modernise" their teaching, who then confuse fixed-response with "modern" assessment and effectively neutralise much of the good they may be achieving in the curriculum.

CORRESPONDENCE: *Alex H. JOHNSTONE, Centre for Science Education, Kelvin Building, University of Glasgow, Glasgow G12 8QQ, UK; tel.: +44 141 330 6565;*
e-mail: alex@chem.gla.ac.uk

APPENDIX: SCORING A STRUCTURAL COMMUNICATION QUESTION

As mentioned earlier in the chapter, this is done in two stages: a score for the selection of pieces of information to answer the question and another score for any sequencing which is required. This can best be shown by an example.

Suppose we have a nine box grid and that the answer to a question is given by the boxes 2, 6, 7 and 9 and the most logical sequence is 6, 7, 2, 9. Now let us score several student attempts.

Student A: Has chosen boxes 2, 7, 8 and 9 and has sequenced them as 8, 7, 2, 9.

$$\text{Choice Score} = 3/4 - 1/5 = 0.75 - 0.2 = 0.55$$

Since there is the possibility of a negative score, this can be eliminated by adding 1 to the choice score before multiplying it by some factor (such as 5) to produce a number which will be recognisable to the student.

Final choice score = (0.55 + 1) x = 1.55 x 5 = 7.75 or rounded up to 8.

The perfect score would have been 10.

Sequence score is based upon a set of yes/no decisions. In this case the most logical order is 6, 7, 2, 9. The questions are:

Does 6 come before 7 (Y/N) and are they adjacent (Y/N)
Does 7 " " 2 (Y/N) " " " " (Y/N)
Does 2 " " 9 (Y/N) " " " " (Y/N)

The student has chosen 8, 7, 2, 9. Now let us apply the test.

Does 6 come before 7 (N) and are they adjacent (N)
Does 7 " " 2 (Y) " " " " (Y)
Does 2 " " 9 (Y) " " " " (Y)

This would score one point for each Y and so would give four marks out of a possible six. The penalty here is because box 6 was omitted in the original choice. The total for the question out of a possible 16 marks is the choice score + sequence score = 8 + 4 = 12.

***Student B**:* Chose 2, 6 and 9 only and sequenced them as 2, 9 and 6.

Choice score = 3/4 - 0/5 = 0.75
Adjusted score = (0.75 + 1) x 5 = 1.75 x 5 = 8.75 = 9.0
Sequence score :
Does 6 come before 7 (N) and are they adjacent (N)
Does 7 " " 2 (N) " " " " (N)
Does 2 " " 9 (Y) " " " " (Y)

This would give a sequence score of 2.

Total score = 9 + 2 = 11 out of a possible 16 marks. This student fared badly on the sequencing and the omission of 7 caused trouble.

Student C: Chose 3, 4, 5, 6 and 7 and sequenced them as 6, 7, 3, 4, 5.

Choice score = 2/4 - 3/5 = 0.5 - 0.6 = -0.1
Adjusted score = (-0.1 + 1) x 5 = 0.9 x 5 = 4.5
Sequence score :
Does 6 came before 7 (Y) and are they adjacent (Y)
Does 7 come before 2 (N) and are they adjacent (N)
Does 2 come before 9 (N) and are they adjacent (N)
The score here is 2.

The total score for the question is 4.5 + 2 = 6.5 (out of 16)

The weighting can be changed in either the choice score or sequence score to achieve the balance the teacher wants. It is wise to inform the students of this balance in advance so that they can share the teacher's view of the relative importance of choice and logical presentation.

REFERENCES

Egan, K. (1972). Structural Communication: A new contribution to pedagogy. *Programmed Learning Educational Technology*, 63-78

Friel, S. & Johnstone, A.H. (1978). Scoring systems which allow for partial knowledge. *Journal of Chemical Education, 55*, 717-719.

Johnstone, A.H. (2001). *Evaluation of teaching.* University of Hull, UK: LTSN Physical Sciences. (ISBN 1-903815-01-0)

Mackenzie, D. (1997). *TRIAD: A computer based assessment software*. University of Derby, UK: Centre for Interactive Assessment Development.

Tamir, P. (1990) Justifying the selection of answers in multiple choice items. *International Journal of Science Education, 12*, 563-573.