Guidelines for Writing Multiple-Choice Items

Derek Cheung
Department of Curriculum and Instruction
The Chinese University of Hong Kong

Introduction

Forty percent of marks scored by physics, biology and chemistry students in the Hong Kong Certificate of Education Examination are assessed through multiple-choice (MC) items. The MC item is also one of the most popular item formats used in classroom assessment. It is unfortunate that most preservice or inservice teacher education programmes offered by tertiary institutions in Hong Kong do not include test construction as a core component. As a result, a graduate may become a qualified teacher without learning any assessment theory or mastering the basic skills in designing assessments. This inadequacy in assessment training can be confirmed by the unsatisfactory quality of the majority of MC items constructed by classroom teachers and textbook publishers. Even public examination papers often contain faulty MC items.[1]

A typical MC item has three parts: a stem that presents a problem; the correct or best answer; and several distractors (i.e., the wrong or less appropriate options). It is common knowledge that the correct answers should be distributed evenly among the alternative positions of MC items, but there are many other important guidelines for writing good items. For example, Haladyna, Downing and Rodriguez (2002) have described 31 guidelines. Space limitation precludes a discussion of all these guidelines here. This paper focuses on eight guidelines that I believe are generally not well recognized by science teachers in Hong Kong.[2] Illustrative examples are provided to demonstrate how these guidelines can be applied to construct chemistry items.

MC Item-Writing Guidelines

1. The stem should be meaningful by itself and should present a definite problem.

A common fault in MC item writing is to have a brief, meaningless stem with problem definition revealed in the options (Ebel & Frisbie, 1991; Haladyna, 1999). In such cases, it can be difficult to see the intent of the item after reading the stem. To write a focused item, we should include the central idea in the stem instead of the options. In Item 1, the stem does not present a definite problem.

ITEM 1

Non-metals

A. cannot exist as solids at room temperature.

B. can combine only with metals to form stable compounds.

*C. usually have more than three electrons in the outermost shell of the atom.

D. are usually found on the left hand side of the Periodic Table.

The correct answer is indicated with an asterisk. Students are faced with four true-false options; each is about non-metals, but only option C is correct. Furthermore, the four options cover a set of widely dissimilar chemical ideas, making the MC item unnecessarily complicated and reducing the diagnostic value of the MC item (Gronlund, 1998). The stem can be judged to be clearly presenting a problem if it forces the options to be parallel in type of content.

Item 2 demonstrates one way to make the stem become a definite problem. Students can think about the correct answer rather than figuring out what the problem is. Also, the clearly stated problem in the stem has forced the four options to be parallel in content.

ITEM 2

How many electrons could be found in the outermost shell of a non-metal atom?

A. 1

B. 2

C. 3

*D. 4

Similarly, Item 3 is a poorly written MC item. The stem fails to present a definite problem and the four options appear to be a hodgepodge of chemical ideas. Clearly, Item 4 is more focused than Item 3. The stem of Item 4 poses a clear, definite problem and assesses a single learning objective.

ITEM 3

Which of the following statements concerning electrochemical cells is correct?

*A. There is a spontaneous chemical reaction in each electrochemical cell.

B. The e.m.f. of an electrochemical cell is measured in joules.

C. The anode is labeled (+) while the cathode is labeled (-).

D. The salt bridge provides electrons to complete the circuit.

ITEM 4

What is the main function of the salt bridge in an electrochemical cell?

*A. supply ions moving to the two half-cells

B. draw electrons from one half-cell to the other half-cell

C. keep the levels of solutions equal in the two half-cells

D. supply electrons to complete the circuit

2. The use of internal or beginning blanks in completion-type MC items should be avoided.

The stem may be written as an incomplete statement that needs to be completed by insertion of the correct option. Measurement specialists have advised not to use the completion format because a student has to retain the stem in short-term memory while completing the stem with each option. Test anxiety is even higher if the student is not a native English speaker. If the completion format is unavoidable, the omission should occur toward the end of the stem rather than in the middle or at the beginning of the stem as shown in Item 5. Item 6 shows an improved version.

ITEM 5

_____have the molecular formula C_nH_{2n}.

A. Alkanes

*B. Alkenes

C. Alkanols

D. Alkanoic acids

ITEM 6

Which type of organic substance has the molecular formula C_nH_{2n}?

A. alkanes

*B. alkenes

C. alkanols

D. alkanoic acids

3. A negatively stated stem is used only when significant learning outcomes require it.

Most students have difficulty understanding the meaning of negatively phrased items. They often read through the negative terms such as *not, no,* and *least*, and forget to reverse the logic of the relation being tested. A positively stated stem also avoids assessing relatively insignificant learning outcomes. Knowing the least important chemical test, the law that does not apply, or the poorest explanation for a chemical change is seldom a significant learning objective. For example, Items 7 and 8 assess the same concept of chemistry, but some students may answer Item 7 incorrectly merely because of the word *least*. Since least and concentrated are opposites, the phrase *least concentrated* is more difficult to understand than the phrase *most concentrated.* Research by Cassels and Johnstone (1984) has confirmed that the change from *least concentrated* to *most concentrated* will increase the percent of correct responses.

ITEM 7

Which of the following solutions is the least concentrated?

A. 50 g of calcium carbonate in 100 cm^3 of water

B. 60 g of sodium chloride in 200 cm^3 of water

C. 65 g of potassium nitrate in 100 cm^3 of water

*D. 120 g of potassium sulphate in 200cm^3 of water

ITEM 8

Which of the following solutions is the most concentrated?

A. 50 g of calcium carbonate in 100 cm^3 of water

B. 60 g of sodium chloride in 200 cm^3 of water

*C. 65 g of potassium nitrate in 100 cm^3 of water

D. 120 g of potassium sulphate in 200 cm^3 of water

Although negatively phrased stems should generally be avoided, they are useful if we want to assess whether students can identify dangerous laboratory practices that may damage expensive equipment or result in bodily injury, and which should not be carried out. Item 9 is an example of such an item. However, when a negative term is used, it should be emphasized by being underlined or capitalized. Replacing the negative term with the word *except* can sometimes improve clarity, as illustrated in Item 10. Few students would overlook the negative element in the stem because the word *except* is deliberately placed at the end of the stem and is capitalized.

ITEM 9

Water-type extinguisher is not suitable for putting out fire caused by burning

*A. alcohol.

B. cotton.

C. paper.

D. wood.

ITEM 10

Water-type extinguisher is suitable for putting out fire caused by burning all of the following EXCEPT

*A. alcohol.

B. cotton.

C. paper.

D. wood.

4. Irrelevant difficulty should be avoided.

The difficulty of an item should not be increased by incorporating more complicated information in the stem than is necessary. For example, if we want to assess whether students can solve dilution problems using the concept of molarity, Item 11 contains confounding detail. The values used in Item 12 will assess the same learning outcome and will avoid irrelevant sources of difficulty and error.

ITEM 11

What volume of water should be added to 57.35 cm^3 of 1.96 M NaCl in order to dilute it to 1.50 M?

*A. 17.59 cm^3

B. 42.65 cm^3

C. 74.94 cm^3

D. 112.41 cm^3

ITEM 12

What volume of water in cubic centimeters should be added to 60 cm^3 of 2.0 M NaCl in order to dilute it to 1.5 M?

*A. 20

B. 40

C. 80

D. 120

Similarly, we should not attempt to increase the difficulty of an item by using unnecessarily complex or unfamiliar vocabulary, such as the word *Topaz* in Item 13. This item aims at testing students' understanding of the types of ions that give a yellow colour. But if students do not know that Topaz is yellow, they are lost. Item 14 is a better measure of the same learning objective. The purpose of chemistry MC tests is to assess students' knowledge, understanding and problem solving, not reading proficiency.

ITEM 13

Which ion below is responsible for the colour of the gemstone called Topaz?

A. Cr^{3+}

B. Cu^{2+}

*C. Fe^{3+}

D. Mn^{3+}

ITEM 14

Which ion below is probably responsible for the colour of yellow gemstones?

A. Cr^{3+}

B. Cu^{2+}

*C. Fe^{3+}

D. Mn^{3+}

5. Plausible distractors should be included.

Designing plausible distractors is the most difficult part of MC item writing (Haladyna, 1999). A good distractor should be selected by low achievers and not by high achievers. To construct plausible distractors, teachers are encouraged to use common misconceptions of chemistry students for distractors. For example, the correct answer in both Items 15 and 16 is pH 3. Many students have the misconception that a solution is drinkable only if its pH is 7 $\pm$ 2 (Cros et al., 1986). Thus, distractors with a pH value within this range are plausible (see Item 16).

ITEM 15

Ordinary soft drinks like Coca-Cola have a pH about

A. 1

B. 2

*C. 3

D. 4

ITEM 16

Ordinary soft drinks like Coca-Cola have a pH about

*A. 3.

B. 5.

C. 6.

D. 8.

For Items 17 and 18, the correct answer is CuS_2. According to Schmidt (1987), students have two common misconceptions. Some tend to use the mass-ratio strategy (Cu:S = 1:1) and select option A in Item 18 as the answer. Other students like to employ the molar-mass-ratio strategy (Cu:S = 64:32) and think that option C in Item 18 is the correct answer. Thus, CuS and Cu_2S are good distractors and useful for diagnosis of students' learning difficulties. Arbitrary distractors such as CuS_3, Cu_2S_3 and Cu_3S should be avoided.

ITEM 17

2 g of a compound contains 1 g copper, the rest is sulphur. Which one of the following formulae correctly represents this compound?

*A. CuS_2

B. CuS_3

C. Cu_2S_3

D. Cu_3S

ITEM 18

2 g of a compound contains 1 g copper, the rest is sulphur. Which one of the following formulae correctly represents this compound?

A. CuS

*B. CuS_2

C. Cu_2S

D. Cu_2S_3

6. The complex MC format should be avoided.

Sometimes, teachers like to design complex MC items to make them harder and to cover more contents. A complex MC item consists of a list of potentially correct answers called primary responses and a list of combinations of the primary responses called secondary options. Students have to select one of the secondary options in answering the item, as shown in Item 19. This item is equivalent to a set of four true-false items, but knowing that a particular primary response is correct or incorrect would help students identify the correct secondary option by eliminating distractors (Ebel & Frisbie, 1991; Haladyna, 1999). Since there are 4 primary responses in Item 19, 15 possible combinations of secondary options are available. Unfortunately, only 4 out of the 15 possible combinations can be included. For example, if a student knows that the primary response "sulphur dioxide" is untrue, he or

she will probably pick option D because sulphur dioxide does not appear in options A and D and more than one primary response are usually included in the correct answer to a complex MC item. Although the complex MC format may make the items more difficult, research (Albanese, 1993; Rodriguez, 1997) has revealed that it is less discriminating and reliable than the single-answer format. Thus, difficult MC items are not necessarily good items. Complex MC items also require more time for students to read, resulting in lower test efficiency (Haladyna, Downing & Rodriguez, 2002; Sax, 1997). Item 20 shows an improved version.

ITEM 19

Which of the following chemicals is/are contained in town gas?

(1) hydrogen

(2) sulphur dioxide

(3) carbon monoxide

(4) gaseous naphtha

A. (3) only

B. (1) and (2) only

C. (2) and (4) only

*D. (1), (3) and (4)

ITEM 20

What is the major constituent of the town gas in Hong Kong?

A. carbon monoxide

B. gaseous naphtha

*C. hydrogen

D. methane

7. The relative length of the options should not provide a clue to the answer.

Teachers are mostly unaware of this item-writing guideline (Rodriguez, 1997). It is common to express the correct response more carefully and at greater length than the distractors. However, research (Chase, 1964) has indicated that longer options tend to result in higher response rates. In Item 21, test-wise students will notice that option D is much longer than the other options. Even without a good understanding of the concept of sacrificial protection, they will guess that the correct answer is D because it stands out from the others. In Item 22, the correct answer is shortened and two distractors are rephrased to the desired length. Although expanding the distractors can increase their specificity and plausibility, teachers should not load them with irrelevant lengthiness or false technicality.

ITEM 21

Why is zinc better than tin if we want to protect a piece of iron from rusting by electroplating?

A. Zinc is cheaper than tin.

B. Tin is toxic.

C. Zinc can prevent iron from contacting

with water and air.

*D. Zinc is more reactive than iron and thus rusting is prevented even when the metal plating is broken.

ITEM 22

Why is zinc better than tin if we want to protect a piece of iron from rusting by electroplating?

A. The cost of extraction of zinc from ores is lower than that of tin.

B. Tin is a toxic metal and causes incurable diseases.

C. Zinc can prevent iron from contacting with water and air.

*D. Rusting is prevented even when the zinc layer is broken.

8. The use of "none of the above" or "all of the above" as an option should be avoided.

The use of *none of the above* and *all of the above* as options in MC items is tempting to many teachers because they appear to fit easily into many items. However, many measurement specialists do not recommend the use of the option *none of the above*. For example, the correct answer for Item 23 is option D. A student may explain this way: "The correct answer is none of the above because, as everyone knows, hydrogen relights a glowing splint." Another student may be surprised to hear that explanation: "What! The correct answer is not hydrogen, but sulphur dioxide." It does not matter; neither gas is listed. Thus, the correct answer could be selected with misinformation. This item may be modified to form Item 24.

ITEM 23

Which of the following substances would relight a glowing splint?

A. carbon dioxide

B. neon

C. nitrogen

*D. none of the above

ITEM 24

Which of the following substances would relight a glowing splint?

A. carbon dioxide

B. neon

C. nitrogen

*D. oxygen

The use of the option *all of the above* is also problematic. For example, Item 25 is poorly constructed because a student may know that two of the three options offered are correct and this information can clue the student into selecting *all of the above*. Thus, the option format allows students to select the correct answer on the basis of only partial rather than complete knowledge of the item. Item 26 shows an improved version.

ITEM 25

Which statement is true of most plastics?

A. They have no reaction with acids.

B. They can be moulded easily.

C. They are flammable.

*D. All of the above.

ITEM 26

Which statement is true of most plastics?

*A. no reaction with acids

B. difficult to be moulded

C. nonflammable

D. good conductors of heat

The option *all of the above* is still faulty even though it is not designed as the correct answer in MC items. When a student recognizes that at least one option is incorrect, he or she may immediately note that the *all of the above* option must be wrong. In such a case, the option *all of the above* is not a functional distractor.

Concluding Remarks

The Hong Kong government has poured millions of dollars into the educational reform with a concern to prepare students for the challenges of a knowledge-based society and globalization. Improving the quality of classroom assessments and public examinations is now one of the foci of our educational reform (Education and Manpower Bureau, 2001). Five basic assessment methods are commonly used in schools: selected response, essay, performance assessment, personal communication, and portfolio. It is important to note that no single method can serve all of our assessment needs; assessing the full range of goals for science education requires applying all of these assessment methods. Selected response assessment includes the MC format. However, poorly written MC items will lower the dependability of test scores. More importantly, poorly written MC items cannot provide science teachers with information to evaluate their instructional effectiveness. The MC format, when properly used, is an efficient and excellent diagnostic tool for identifying science students' misconceptions. In this paper, I discuss eight useful guidelines for writing MC items and demonstrate their appropriate use with chemistry items. They are based on consensus of measurement specialists, and some of them have been supported by empirical research (Haladyna, Downing & Rodriguez, 2002). To construct high-quality MC items, Hong Kong science teachers, textbook writers and setters of public examinations must have a deep understanding of these item-writing guidelines.

Note:

1.To take a look at some faulty MC items used in the Hong Kong Certificate of Education Examinations, science teachers may browse the CD-ROM "Test Construction Support System for Chemistry Teachers (version 2.0)" distributed by the Chinese University of Hong Kong.

2.An earlier version of this paper was presented at the Science and Technology Education Conference held in Hong Kong on 21 June 2002.

References

Albanese, M. A. (1993). Type K and other complex multiple-choice items: An analysis of research and item properties. *Educational Measurement: Issues and Practice,* 12, 28-33.

Cassels, J. R. T., & Johnstone, A. H. (1984). The effect of language on student performance on multiple choice tests in chemistry. *Journal of Chemical Education, 61*(7), 613-615.

Chase, C. I. (1964). Relative length of option and response set in multiple choice items. *Educational and Psychological Measurement,* 24(4), 861-866.

Cros, D., Maurin, M., Amouroux, R., Chastrette, M., Leber, J. & Fayol, M. (1986). Conceptions of first-year university students of the constituents of matter and the notions of acids and bases. *European Journal of Science Education,* 8(3), 305-313.

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement.* Englewood Cliffs, NJ: Prentice Hall.

Education and Manpower Bureau (2001). *Reform of the education system in Hong Kong: Summary.* Hong Kong: Printing Department.

Gronlund, N. E. (1998). *Assessment of student achievement.* Boston: Allyn and Bacon.

Haladyna, T. M. (1999). *Developing and validating multiple-choice test items.* Mahwah: Lawrence Erlbaum.

Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education,* 15(3), 309-334.

Rodriguez, M. C. (1997). *The art & science of item writing: A meta-analysis of multiple-choice item format effects.* Paper presented at the annual meeting of the American Educational Research Association.

Sax, G. (1997). *Principles of educational and psychological measurement and evaluation.* Belmont, CA: Wadsworth.

Schmidt, H. J. (1987). Secondary school students' learning difficulties in stoichiometry. In J. Novak (Ed), *Misconceptions & educational strategies in science and mathematics* (pp. 396-404). Ithaca: Cornell University.