October 2005
ISSN 1109-4028

Volume 6, Issue No 4
Pages 180 – 212

Chemistry Education Research and Practice

Published quarterly by The Royal Society of Chemistry

RS•C
ROYAL SOCIETY OF CHEMISTRY

Chemistry Education Research and Practice

October 2005
ISSN 1109-4028

Volume 6, Issue No 4
Pages 180-212

Contents

Papers

Indexed/Abstracted in
CHEMICAL ABSTRACTS (CA)
EDUCATIONAL RESEARCH ABSTRACTS ONLINE (ERA)
http://www.tandf.co.uk/era

Chemistry Education Research and Practice

The journals, *University Chemistry Education*, published by The Royal Society of Chemistry, (http://www.rsc.org/uchemed/uchemed.htm) and *Chemistry Education Research and Practice*, published from the University of Ioannina, (http://www.uoi.gr/cerp/) have merged with effect from January 1st 2005. The new, fully electronic journal is published by The Royal Society of Chemistry under the title: ***Chemistry Education Research and Practice,*** and it will continue to be available free of charge on the Internet**.** There are four issues per year.

The new journal is edited by Georgios Tsaparlis (gtseper@cc.uoi.gr) and Stephen Breuer (s.breuer@lancaster.ac.uk) and intends to maintain the high standards set by its predecessors. Its editorial policy will be the following.

'***Chemistry Education Research and Practice***' is the journal for teachers, researchers and other practitioners in chemical education. It is the place to publish papers on:
- research, and reviews of research in chemical education;
- effective practice in the teaching of chemistry;
- in depth analyses of issues of direct relevance to chemical education

Contributions can take the form of full papers, preliminary communications, perspectives on methodological and other issues of research and/or practice, reviews, letters relating to articles published and other issues, and brief reports on new and original approaches to the teaching of a specific topic or concept.

The new journal welcomes contributions of the type described above; these should be sent to cerp@rsc.org.

Chemistry Education Research and Practice

Chemistry Education Research and Practice

Guidelines for Authors

Submission of contributions

Chemistry Education Research and Practice (CERP) is the journal for teachers, researchers and other practitioners in chemical education. It is published free of charge, electronically, by The Royal Society of Chemistry, four times a year. It is the place to publish papers on:

- research, and reviews of research in chemical education;
- effective practice in the teaching of chemistry;
- in depth analyses of issues of direct relevance to chemical education

Contributions can take the form of full papers, preliminary communications, perspectives on methodological and other issues of research and/or practice, reviews, letters relating to articles published and other issues, and brief reports on new and original approaches to the teaching of a specific topic or concept.

1. The original contribution should be submitted electronically, preferably in Word for Windows format. Any associated diagrams should be attached in JPG or GIF format, if possible. Submissions should be made by e-mail as a file attachment to cerp@rsc.org, or directly to the editors: Stephen Breuer at s.breuer@lancaster.ac.uk or to Georgios Tsaparlis (gtseper@cc.uoi.gr).

2. Submitted contributions are expected to fall into one of several categories (listed above). Authors are invited to suggest the category into which the work should best fit, but the editors reserve the right to assign it to a different category if that seems appropriate.

   A word count (excluding references, tables, legends etc) should be included at the end of the document.

3. Presentation should be uniform throughout the article.

   Text should be typed in 12pt Times New Roman (or similar), with 1"/ 2.5 cm margins, double-spaced, unjustified, ranged left and not hyphenated.

   Always use an appropriate mix of upper and lower case letters: do not type words in uppercase letters either in the text or in headings. **Bold** or *italic* text and not upper case letters should be used for emphasis.

   All nomenclature and units should comply with IUPAC conventions.

   Tables and figures should be numbered consecutively as they are referred to in the text (use a separate sequence of numbers for tables and for figures). Each should have an informative title and may have a legend.

Equations should be written into the text using the word processing program, either as normal text or using the program's equation facility.

Structures should, wherever possible, be treated as a figure and not incorporated into text.

References should be given by the name of the author (or the first author, if more than one), followed by the year of publication. If an author has more than one reference from the same year, then it should be given as Smith 2001a, Smith 2001b, etc.

Footnotes should be generally avoided and important additional information may be referenced and included in the reference list.

4. A title page must be provided, comprising:
   - an informative title;
   - authors' names and affiliation, full postal address and e-mail; (in the case of multi-authored papers, use an asterisk to indicate one author for correspondence, and superscript a, b, etc. to indicate the associated addresses);
   - an abstract of not more than 200 words;
   - keywords identifying the main topics covered in the paper

5. Wherever possible articles should be subsectioned with headings, subheadings and sub-sub-headings. Do **not** go lower than sub-sub-headings. Sections should not be numbered.

   The introduction should set the context for the work to be described; include references to previous related work, and outline the educational objectives.

   A concluding section (which need not be headed conclusion) will include an evaluation of the extent to which educational objectives have been met. A subjective evaluation may be acceptable.

6. The formatting of references should follow the following practice:

   Books and Special Publications:
   Author A., (year), *Title of the book italicized*, Publisher, Place of publication, page no. if applicable.

   Journal Articles:
   Author A., Author B. and Author C., (year), Title of the article in Roman type, *Full Name of the Journal Italicised*, **Volume no. in Bold**, inclusive page numbers.

   For example:
   Osborne R. and Freyberg P., (1985), *Learning in science: the implication of children's science*, Heinemann, London.

   Jackman L.E. and Moellenberg W., (1987), Evaluation of three instructional methods for teaching general chemistry, *Journal of Chemical Education*, **64**, 794-96.

7. All contributions submitted will be refereed anonymously by two independent referees. In case of a disagreement a third referee will be consulted. The decision of the Editors on

the acceptance of articles is final.

8. Authors grant *CERP* the exclusive right to publish articles. They undertake that their article is their original work, and does not infringe the copyright of any other person, or otherwise break any obligation to, or interfere with the rights of such a person, and that it contains nothing defamatory.

9. Articles will be published on the Web in PDF and HTML formats.

Educational Research

The ionisation energy diagnostic instrument: a two-tier multiple-choice instrument to determine high school students’ understanding of ionisation energy

Kim-Chwee Daniel Tan*[a], Keith S. Taber[b], Ngoh-Khang Goh[a] and Lian-Sai Chia[a]

[a] *National Institute of Education, Nanyang Technological University, Singapore*
e-mail: kcdtan@nie.edu.sg
[b] *University of Cambridge Faculty of Education*
e-mail: kst24@cam.ac.uk

Received 28 June 2005, accepted 15 September 2005

Abstract: The topic of ionisation energy is important as the concepts involved provide the foundation for the understanding of atomic structure, periodic trends and energetics of reactions. Previous research has shown that A-level (high school) students in the United Kingdom had difficulty understanding the concepts involved in ionisation energy. This paper describes the development and administration of a two-tier, multiple-choice instrument on ionisation energy, the Ionisation Energy Diagnostic Instrument, to determine if A-level students (Grade 11 and 12, 17 to 18 years old) in Singapore have similar alternative conceptions to those of their counterparts in the United Kingdom, as well as explore their understanding of the trend of ionisation energies across Period 3. The items in such instruments are specifically designed to identify alternative conceptions and misunderstandings in a limited and clearly defined content area. The results showed that students in Singapore applied the same octet rule framework and conservation of force thinking to explain the factors influencing ionisation energy as the students in the United Kingdom. In addition to the above alternative frameworks, many students in Singapore also resorted to relation-based reasoning to explain the trend of ionisation energies across Period 3 elements. [*Chem. Educ. Res. Pract.*, 2005, **6** (4), 180-197]

Keywords: alternative conceptions in chemistry, assessment, high school chemistry, ionisation energy, physical chemistry, two-tier multiple-choice diagnostic test.

Introduction

Many researchers agree that the most important significant things that students bring to class are their conceptions (Ausubel, 1968, 2000; Driver et al., 1986). Duit and Treagust (1995) define conceptions as “*the individual’s idiosyncratic mental representations*”, while concepts are “*something firmly defined or widely accepted*” (p. 47). Children develop ideas and beliefs about the natural world through their everyday life experiences. These include sensual experiences, language experiences, cultural background, peer groups, mass media, as well as formal instruction (Duit et al., 1995). Studies have revealed that students bring with them to science lessons certain ideas, notions and explanations of natural phenomena that are inconsistent with the ideas accepted by the scientific community (Osborne et al., 1983). These existing ideas are often strongly held, resistant to traditional teaching and form coherent though mistaken conceptual structures (Driver et al., 1978). Students may undergo instruction in a particular science topic, do reasonably well in a test on the topic, and yet, do not change their original ideas pertaining to the topic even if these ideas are in conflict with the scientific concepts they were taught (Fetherstonhaugh et al., 1992). Duit and Treagust

(1995) attribute this to students being satisfied with their own conceptions and therefore seeing little value in the new concepts. Another reason they proposed was that students look at the new learning material "*through the lenses of their preinstructional conceptions*" (p. 47) and may find it incomprehensible. Osborne et al. (1983) state that students often misinterpret, modify or reject scientific viewpoints on the basis of the way they really think about how and why things behave, so it is not surprising that research shows that students may persist almost totally with their existing views (Treagust et al., 1996). In this paper, the term 'alternative conceptions' is used to describe student conceptions that differ from scientific concepts. The authors agree with Wandersee, Mintzes and Novak (1994) that the term "*confers intellectual respect on the learner who holds those ideas – because it implies that alternative conceptions are contextually valid and rational and can lead to even more fruitful conceptions (e.g., scientific conceptions)*" (p. 178).

Thus, students' alternative conceptions have to be identified so that measures can be taken to help students replace them with (or develop them into) more scientifically acceptable concepts (Taber, 1998a). Studies in which students' alternative conceptions are described cover a wide range of subject areas including chemistry (Garnett et al., 1995; Barker et al., 2000; Pedrosa et al., 2000; Schmidt, 2000; Taber et al., 2000; Taber, 2001; De Jong et al., 2002; Harrison et al., 2002). A useful review of alternative conceptions at secondary school level was provided by Driver et al. (1994). Besides exploring and identifying students' alternative conceptions, most of these studies provide implications for the teaching and learning of the concepts examined.

Methods used to determine students' understanding of concepts include concept mapping (Novak, 1996), interviews (Carr, 1996) and multiple-choice diagnostic instruments (Treagust, 1995). In this study, a two-tier multiple-choice instrument (Treagust, 1995) was developed and used to determine students' understanding of the concepts involved in ionisation energy. The items in two-tier multiple-choice diagnostic instruments are specifically designed to identify alternative conceptions and misunderstandings in a limited and clearly defined content area. The first part of each item consists of a multiple-choice content question having usually two or three choices. The second part of each item contains a set of four or five possible reasons for the answer to the first part; this makes the diagnostic instrument more powerful and effective as it allows an insight to the underlying reasons for the student's answer. Incorrect reasons (distracters) are derived from actual student alternative conceptions gathered from the literature, interviews and free response tests. This methodology has been used to develop diagnostic tests in chemistry, for example, on covalent bonding (Peterson et al., 1989), chemical bonding (Tan et al., 1999), chemical equilibrium (Tyson et al., 1999), and qualitative analysis (Tan et al., 2002).

The topic of ionisation energy has traditionally formed part of the A-level chemistry curriculum studies in many parts of the world, including Singapore and the UK. The topic is important as the concepts involved also provide the foundation for the understanding of atomic structure, periodic trends and energetics of reactions (Taber, 2003a). However, it is a difficult topic to learn because it involves "*abstract and formal explanations of invisible interactions between particles at a molecular level*" (Carr, 1984, p. 97). For example, Taber (1998a,b, 1999, 2003a) found that students had difficulty in understanding the principles determining the magnitude of ionisation energy. This is because students based their explanations on the octet rule/full shell framework and 'conservation of force' conception, and did not or could not apply basic electrostatic principles that they learned in physics to explain the interactions between the nucleus and electrons in an atom.

Purpose

The study sought to develop a two-tier multiple-choice diagnostic instrument to determine if A-level students in Singapore had similar alternative conceptions and explanatory principles of the factors influencing ionisation energy as their A-level counterparts in the United Kingdom, as well as to explore students' conceptions of the trend of ionisation energies across different elements in the Periodic Table. This extension of Taber's (1999) study to include students' conceptions of the trend of ionisation energies was in line with requirements of the A-level chemistry syllabus on ionisation energy (Appendix A). By knowing their students' alternative conceptions of ionisation energy, teachers can gain a greater insight into the subject matter of the topic, their teaching, and the learning processes of their students. They are also likely to be more receptive and willing to try or develop alternative teaching strategies if they find that their present methods are inadequate in addressing students' difficulties.

Method and procedures

The two-tier multiple-choice diagnostic instrument on ionisation energy was developed in three phases using procedures defined by Treagust (1995). The first phase involved the first-named author defining the content framework of A-level ionisation energy with a concept map (Figure 1) and a list of propositional knowledge statements based on Taber's (1997a, 1999) work, an extract of the sections of the A-level chemistry syllabus relevant to ionisation energy (Appendix A), and two chemistry textbooks. The concept map and propositional knowledge were reviewed by 13 experienced A-level chemistry teachers and two tertiary chemistry educators. The reviewers agreed that the concept map and propositional knowledge statements met the requirements of the A-level chemistry syllabus on ionisation energy (Appendix A) in terms of accuracy and relevance. The assessment of the mastery of the content would then be administered in accordance to this framework and this would ensure the content validity of the assessment.

Phase Two was carried out in three stages. In Stage 1, a justification multiple-choice instrument in which students had to supply reasons for their choice of options was developed, based on the propositional knowledge statements on ionisation energy and the findings of Taber's (1999) research. The items in the instrument tested students' understanding of the factors influencing ionisation energy as well as the trend of ionisation energies across a period. This instrument was administered to eighteen Grade 11 students after they were taught ionisation energy. Another six Grade 11 students were interviewed in pairs using the instrument as the interview protocol. The results obtained led to the development of the second version of the justification multiple-choice instrument, which was administered to 146 Grade 11 and 12 students from four schools in Stage 2. The third version of the justification multiple-choice instrument, whose development was based on the results obtained in Stage 2, was administered to 130 Grade 12 students from three schools in Stage 3. Eleven Grade 12 students who took the test were interviewed using the instrument as the interview protocol to determine whether any item was ambiguous and to probe the reasons for their answers. The data collected from Phase Two of the study was reported elsewhere (see Tan et al., 2003).

The results from the administration of the justification multiple-choice instruments and the interviews with students in Phase Two contributed to the development of the first version of the two-tier multiple-choice instrument in Phase Three. Further trials and refinement involving 283 Grade 11 and 12 students led to the development of the second version, and subsequently, the final version of the diagnostic instrument, the Ionisation Energy Diagnostic Instrument (IEDI), presented in Appendix B.

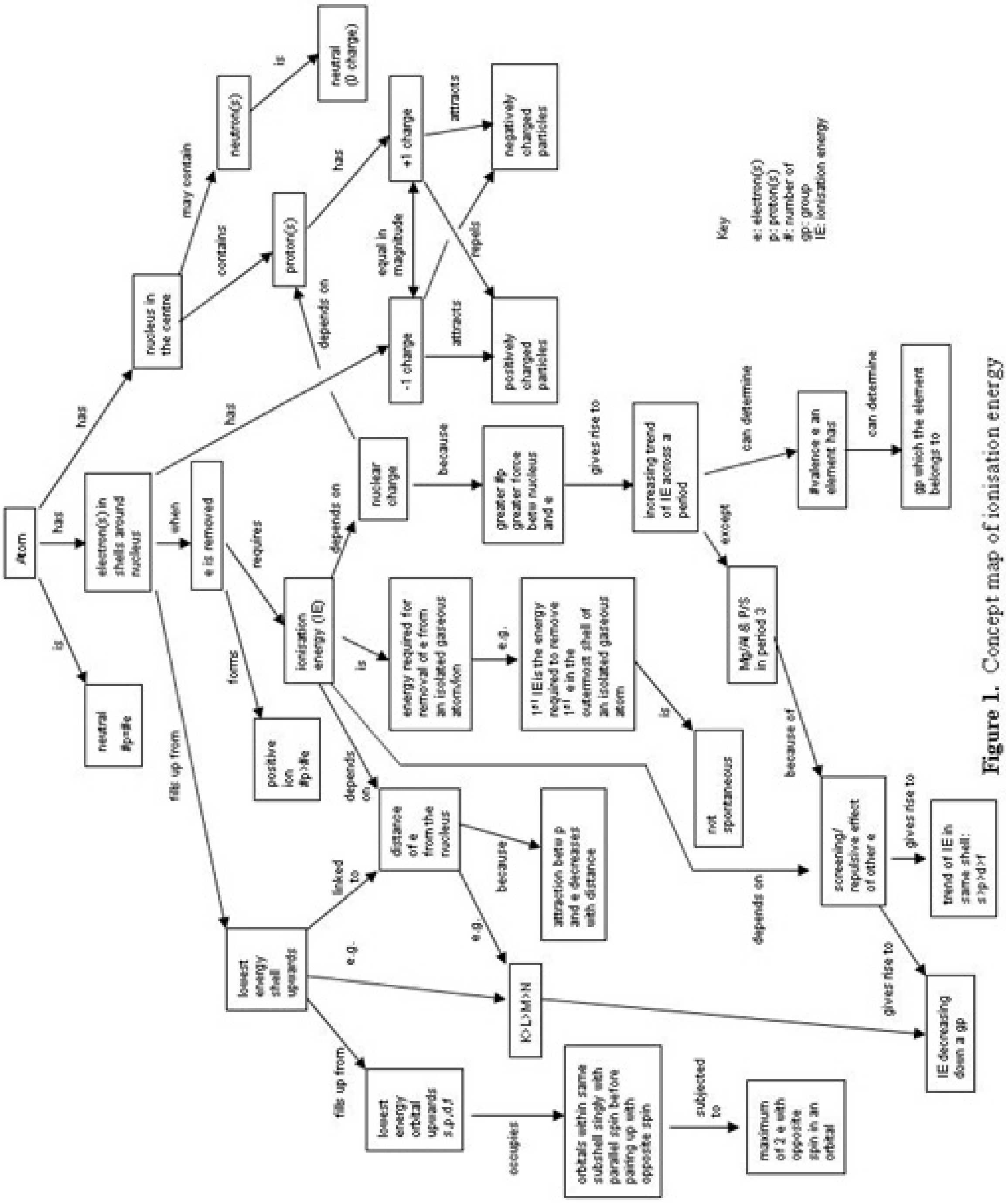


Figure 1. Concept map of ionisation energy

The IEDI was validated by five experienced A-level chemistry teachers and two tertiary chemistry educators for accuracy and relevance. The IEDI was administered to 777 Grade 11 and 202 Grade 12 students from eight out of a total of seventeen A-level institutions in Singapore in June and July 2003. Thirty-two Grade 11 and 12 students, selected by their teachers, were interviewed, either in pairs or in groups of four, using the IEDI as the protocol to determine if there was any ambiguity in the items and to further probe the thinking behind their answers. Each interview lasted between 40 minutes to an hour, and was transcribed verbatim.

Results

Table 1 describes the percentage of the Grade 11 and 12 students selecting each response combination for each item in the IEDI. The results for an item will not add up to 100% if there were students who did not select a response to both parts of the item, selected an answer combination which was beyond the options given in the item, or selected more than one answer combination.

Table 1. The percentage of Grade 11 and 12 students (n=979) selecting each response combination for each item in the IEDI

Item	Content option	Reason option					Total
		(1)	(2)	(3)	(4)	(5)	
1	A	4.8	*43.6*	3.3	-	-	51.7
	B	1.0	5.3	**38.2***	-	-	44.5
	C	0	.2	.2	-	-	0.4
2	A	6.5	.7	*49.7*	-	-	56.9
	B	**30.0***	1.4	3.7	-	-	35.1
	C	.4	.2	.4	-	-	1.0
3	A	**16.8***	.9	4.5	3.8	-	26.0
	B	1.7	1.3	2.2	*63.6*	-	68.8
	C	.1	0	.1	.4	-	0.6
4	A	*15.6*	*18.0*	**48.1***	3.6	-	85.3
	B	.1	.3	.7	1.6	-	2.7
	C	5.3	.3	.1	0	-	5.7
	D	.1	.1	0	0	-	0.2
5	A	1.2	2.2	2.9	*22.0*	4.2	32.5
	B	*13.1*	9.1	**29.2***	9.3	2.2	62.9
	C	.1	.1	.2	0	0	0.4
6	A	6.2	*48.1*	2.9	**5.4***	5.5	68.1
	B	.9	7.2	8.5	1.8	8.3	26.7
	C	.1	.1	.1	.1	.1	0.5
7	A	.5	1.8	*20.7*	*24.4*	1.5	48.9
	B	.7	5.6	7.4	6.8	**23.9***	44.4
	C	.1	0	.2	.1	.2	0.6
8	A	3.9	4.5	7.6	5.8	-	21.8
	B	5.5	*24.9*	4.6	**34.0***	-	69.0
	C	.3	.1	.1	.2	-	0.7
9	A	2.7	3.9	*19.6*	7.4	**32.1***	65.7
	B	6.8	1.6	3.5	*10.4*	4.3	26.6
	C	.2	.1	.3	.4	0	1.0
10	A	6.8	6.8	6.3	*19.0*	-	38.9
	B	3.6	3.5	**33.1***	9.0	-	49.2
	C	.3	.1	.6	.3	-	1.3

Note: Figure in bold and with an asterisk indicates the correct answer.
Figure in italics indicate a major alternative conception (>10%)

Alternative conceptions

Alternative conceptions are considered significant and common if they were found in at least 10% of the student sample (Peterson, 1986). If a higher minimum value, say 25%, was chosen, this would mean not discussing alternative conceptions that seemed likely to be found among students in many classes. Table 2 summarises the significant common alternative conceptions determined from the administration of the IEDI to the 979 A-level students. Eleven significant common alternative conceptions were identified and grouped under the headings of 'Octet rule framework', 'Stable fully-filled or half-filled sub-shells', 'Conservation of force thinking' and 'Relation-based reasoning'.

Table 2. Alternative conceptions determined from the administration of the IEDI

Alternative conception	Choice combination	Percentage of students with the alternative conception
Octet rule framework		
The sodium ion will not recombine with an electron to reform the sodium atom, as its stable octet configuration would be disrupted.	Q1 (A2)	44
The Na(g) atom is a less stable system than the $Na^+(g)$ and a free electron because the $Na^+(g)$ has a stable octet configuration.	Q3 (B4)	64
The second ionisation energy of sodium is higher than its first because the stable octet would be disrupted.	Q4 (A1)	16
Stable fully-filled or half-filled sub-shells		
The first ionisation energy of sodium is less than that of magnesium because magnesium has a fully filled 3s sub-shell.	Q5 (B1)	13
The first ionisation energy of silicon is less than that of phosphorus because the 3p sub-shell of phosphorus is half-filled.	Q8 (B2)	25
The first ionisation energy of phosphorus is greater than that of sulfur because the 3p sub-shell of phosphorus is half-filled, hence it is stable.	Q9 (A3)	20
Conservation of force thinking		
When an electron is removed from the sodium atom, the attraction of the nucleus for the 'lost' electron will be redistributed among the remaining electrons in the sodium ion.	Q2 (A3)	50
The second ionisation energy of sodium is greater than its first ionisation energy because the same number of protons in the Na^+ ion attracts one less electron, so the attraction for the remaining electrons is stronger.	Q4 (A2)	18
Relation-based reasoning		
The first ionisation energy of magnesium is greater than that of aluminium because the 3p electron of aluminium is further from the nucleus compared to the 3s electrons of magnesium.	Q6 (A2)	48
The first ionisation energy of sodium is greater than that of aluminium because the 3p electron of aluminium experiences greater shielding from the nucleus compared to the 3s electron of sodium.	Q7 (A3)	21
The first ionisation energy of sodium is greater than that of aluminium because the 3p electron of aluminium is further away from the nucleus compared to the 3s electron of sodium.	Q7 (A4)	24

Options A4 of item 5, B4 of item 9, and A4 of item 10 (see Table 3) were not considered as alternative conceptions even though they were incorrect. These questions dealt with the trend of ionisation energies across Period 3. In the items, students had to consider which important factors were in play, as well as to decide which factor outweighed the other (nuclear attraction versus electron shielding/repulsion) in the specific instance. If a student chose one of the stated options, it could indicate that he/she knew which two factors were in play, but decided wrongly on the more important factor in that specific situation. Thus, it was difficult to determine if the student had an alternative conception, or if he/she forgot or could not decide which factor outweighed the other in that specific situation. In other words, these errors are better considered failures of recall than lack of understanding of the concepts involved.

Table 3. Significant errors of students (10% or greater), which were not considered as alternative conceptions

Item	Option	Errors	Percentage of Grade 11 and 12 students
5	A4	The first ionisation energy of sodium is greater than that of magnesium because the paired electrons in the 3s orbital of magnesium experience repulsion from each other and this effect is greater than the increase in the nuclear charge in magnesium.	22
9	B4	The first ionisation energy of phosphorus is less than that of sulfur because the effect of an increase in nuclear charge in sulfur is greater than the repulsion between its 3p electrons.	10
10	A4	The first ionisation energy of silicon is greater than that of sulfur because the effect of an increase in nuclear charge in sulfur is less than the repulsion between its 3p electrons.	19

Octet rule framework

Many students (44%) thought that the sodium ion would not recombine with an electron to reform the sodium atom because the sodium ion had already achieved a noble gas configuration, and gaining an electron would cause the ion to lose its stability (Item 1, A2). In item 3, 64% agreed that the 'sodium ion and a free electron' system was more stable than the sodium atom because the outermost shell of the ion had achieved a stable octet/noble gas configuration (B4). When asked during interviews why an octet configuration gave the sodium ion 'stability', several students either stated that they were taught so, or that it was because the outermost shell of the sodium ion was filled so it could not gain or lose any electrons.

I: Why will the ion be more stable?
P14: Because octet…stable octet configuration.
I: Why is an octet configuration stable?
P14: Because it has eight electrons in the outermost shell already…so no electrons will go in, no electrons will go out…then it is very stable…to achieve a stable configuration…stable octet structure you need to have all your…shells filled… outermost shell filled.

(I represents the Interviewer, P14 represents Pupil 14.)

Cross-tabulation was used to study the consistency of the students' answers (Tan et al., 2002). Cross-tabulation of items 1 and 3 showed that only ninety students (9%) had both items correct, and 323 students (34%) consistently used the octet rule framework in both items (item 1: A2, 44%; item 3: B4, 64%). In addition, 211 students who had item 1 correct used the octet rule framework in item 3. This indicated that students could have and use both the correct concepts and the octet rule framework, even if this resulted in conflicting answers in different items. The explicit comparison of the stability of the sodium atom with the system consisting of the sodium ion and free electron could have influenced the students' use of the octet rule framework in item 3.

Students also used the octet rule framework to justify why the second ionisation energy of sodium was greater than its first ionisation energy (Item 4, A1, 16%). This differs from the curriculum model, which states that the removal of the second electron from sodium involves removing an electron from an inner (second) shell which is more strongly attracted to the nucleus as it is closer to the nucleus and experiences shielding/screening from only two electrons in the first shell. Alternatively, this last factor may be described in terms of the core charge (Taber, 2002; 2003a).

Cross-tabulation of items 1, 3 and 4 showed that only sixty-two students (6%) consistently adopted the octet rule framework in all three items (item 1: A2, 44%; item 3: B4, 64%; item 4: A1, 16%). One hundred sixty-one students (16%) who had item 4 correct used the octet rule framework in items 1 (A2, 44%) and 3 (B4, 64%), and fifty-two students (5%) who adopted the octet framework in items 1 and 3 used the conservation of force thinking (to be discussed in a later section) in item 4 (A2, 18%). The lack of consistency of alternative conceptions held by the students could point to students having more than one conception for a particular concept and "*different conceptions can be brought into play in response to different problem contexts*" (Palmer, 1999, p. 639). Taber (1999, 2000) and Voska et al. (2000) also found that students gave inconsistent answers to apparently related items. One interpretation is that students may be in the process of transition, for example, between holding an alternative conception and adopting the approved curriculum model (Taber, 2000). Caravita et al. (1994) talk about the 'meta-level' of cognitive structure where metacognitive and epistemological beliefs may influence the conceptions that are accessed and applied. Finally, the lack of consistency could also be due to students not having adequate understanding of the topic and resorting to guessing.

Stable fully-filled or half-filled sub-shells

In item 5, 13% stated that magnesium had a higher first ionisation energy than sodium because magnesium had a fully-filled 3s orbital/sub-shell which gave it stability (B1), while in items 8 and 9, 25% (B2) and 20% (A3), respectively, indicated that phosphorus had a higher first ionisation energy compared to silicon and sulfur because the 3p sub-shell of phosphorus was half-filled, hence it was stable. The excerpt of an interview below illustrates this 'stable fully-filled or half-filled sub-shell' thinking:

P14: I put B1 (item 5)…because the magnesium…the last orbital…the 3s orbital is fully filled so it will tend to be more stable…and when an orbital is either half or fully filled it will be more stable…so since sodium has only one electron in the…so when fully filled will be more stable…sodium has only one electron in the s orbital…so to be more stable it will tend…it will be easier to remove…that electron and so the ionisation energy will be lower than that of magnesium.
I: So you are saying that the first ionisation energy of magnesium is higher?
P14: It is more stable.
I: Because of the…
P14: It's fully filled orbital.

I: So why is this fully filled orbital stable…the $3s^2$?
P14: Just like in the shell…I mean…to achieve the octet structure…must have eight electrons…so when you have eight electrons in the outer shell…will be more stable… so when the orbital is fully filled then… more stable… because there's…like…more stable.
I: So stability comes with filled orbitals?
P14: Fully filled and half filled…but fully filled will have higher stability.

P14 believed that a fully-filled or half-filled sub-shell was stable because both were analogous to a 'stable octet' – there was no conflict between the octet rule framework and the 'stable fully-filled or half-filled sub-shell' thinking, in fact, the former seemed to lead 'naturally' to the latter, from shell to sub-shell. In the curriculum model, magnesium has a higher first ionisation energy than sodium because its greater nuclear charge outweighs the repulsion between its 3s electrons. A similar reason accounts for the higher first ionisation energy of phosphorus compared to silicon. However, sulfur has lower first ionisation energy than phosphorus even though sulfur has a greater nuclear charge. This is because the repulsion between the paired 3p electrons in sulfur outweighs its greater nuclear charge. The greater shielding of the 3p electron by the inner shell electrons as well as the 3s electrons explains why aluminium has a lower first ionisation energy compared to magnesium. It has to be noted that students will have great difficulty answering questions on ionisation energy trends if they cannot *either* remember whether increased nuclear charge or increased repulsion/shielding between electrons is more important in specific cases *or* recall the shape of the trend graph, and so work out which factors must be more important in each case. Thus, as mentioned earlier, it is not a matter of grave concern if students cannot decide between, for example, A4 or B3 in item 5, or A5 or B4 in item 9. However, it is problematic when students think that a fully-filled 3s sub-shell gives magnesium its stability, and hence higher first ionisation energy compared to aluminum, while phosphorus, with its 3p sub-shell half-filled, is more stable than sulfur and hence has higher first ionisation energy than sulfur.

Conservation of force thinking

Students indicated in item 2 (A3, 50%) that the nuclear attraction would be redistributed among the remaining 10 electrons when an atom of sodium loses an electron because the number of protons was the same but there was one less electron to attract. The curriculum model states that the net attraction for an electron by the nucleus depends on the number of protons in the nucleus, the distance of the electron from the nucleus and the shielding effect of other electrons in the atom. Removal of one electron from the sodium atom may reduce some repulsion between electrons causing the remaining 10 electrons to move closer to the nucleus, but the nuclear attraction for the electron which was removed is not redistributed to the remaining 10 electrons. Though conceptually incorrect, the conservation of force thinking "*does often allow correct predictions to be made (successive ionisation energies do increase) and seems to have an intuitive attraction to many students*" (Taber, 2003a, p. 156). This was shown in item 4 (A2) where 18% thought that the second ionisation energy of sodium was greater than its first because the same number of protons in sodium was attracting 10 electrons now instead of 11. The excerpt of an interview below illustrates the conservation of force thinking:

P4: Ok…I think it is true (item 2) because…like one electron is lost…the atom has one electron less, right…so…the attraction will just remain the same…so the other electrons have …greater attraction.
I: So you believe the electron…
P4: The attraction stays the same…when the electron goes out, the attraction doesn't go with the electron…so the other electrons experience greater attraction.

I: So it experiences the attraction left behind by the electron…OK P3 what is your reason?
P3: It's the same…since one electron is removed right…so the proton number is the same…so the protons has lesser electrons to attract…so the attraction force is greater …it will pull [the electrons] closer.

Cross-tabulation showed that 134 students (14%) consistently exhibited the conservation of force thinking in items 2 (A3, 50%) and 4 (A2, 18%). This indicated that students who chose the conservation of force option in item 4 were also likely to choose the similar option in item 2. Students can hold both the correct concept and alternative conception as shown by 215 students (22%) who had item 4 correct (A3, 48%) but chose the conservation of force option in item 2 (A3, 50%), and the written answer of one student, "*The same number of protons in* Na^+ *attracts one less electron, so the attraction for the remaining electron is stronger, moreover the second electron is located nearer to the nucleus*". There will be no cognitive conflict deriving from holding alternative conceptions like these where the conservation of force thinking and the curriculum model (as discussed above) lead to the same outcomes – in this case a greater value for the ionisation energy. Students could also hold more than one alternative framework – fifty-two students (5%) who adopted the octet rule framework options in items 1 (A2, 44%) and 3 (B4, 64%) used the conservation of force thinking in item 4 (A2, 18%), while sixty-seven students (7%) who chose the conservation of force option in item 2 (A3, 50%) selected the octet rule framework option in item 4 (A1, 16%).

Relation-based reasoning

Factors influencing ionisation energy include the nuclear charge, the distance of the electron from the nucleus and the repulsion/screening effect of the other electrons present. The results from items 5 to 10 on the trend of the first ionisation energy across Period 3 showed that many students did not consider all the three factors but based their reasons exclusively on one or two factors. Driver et al. (1996) describe this type of thinking as relation-based reasoning, where "*students tend to consider only one factor as possibly influencing the situation – the one which they see as the 'cause'*"(p. 115), and thus, overlook other possible influential factors. For example, many students indicated that the first ionisation energies of magnesium and sodium were greater than that of aluminum because the 3p electron of aluminum was further away from the nucleus than the 3s electron(s) of magnesium (item 6: A2, 48%) and sodium (item 7, A4, 24%), respectively. However, in the curriculum model, atomic radii decrease from sodium to sulfur in Period 3 because of increasing nuclear charge, which outweighs the increase in repulsion between the increasing number of electrons in the same shell.

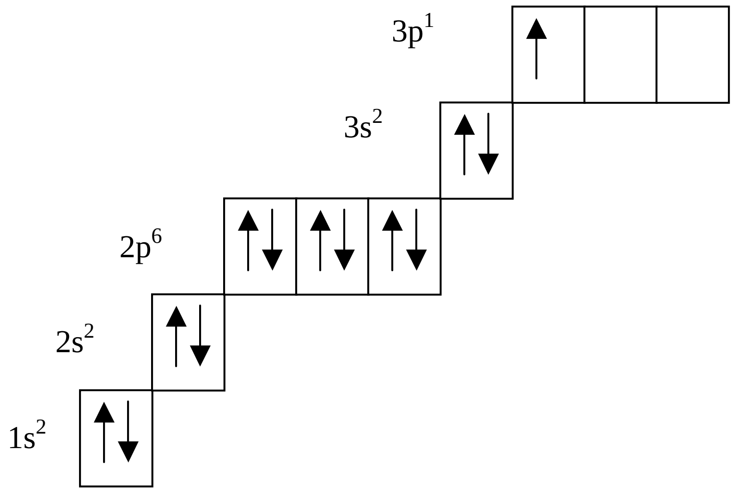
P15: I put A2 (item 6)…because the 3p electron of aluminium is further, right…so they will be further from the nucleus…because they experience…the attraction won't be so strong.
I: What made you say that the 3p electron of aluminium is further from the nucleus compared to the 3s of magnesium?
P13: We were taught that way.
P15: It's further…is taught it's further…it's taught during lectures.
P14: 1s, 2s, 2p, 3s, further, further, further.
P15: Further away.
P14: The 3p is further away.

Note that here students are remembering a comparison presented in the context of a single atomic system (e.g. 3s of sodium compared with 3p of sodium), and expecting the same pattern when the comparison is made between different systems (e.g. 3s of magnesium compared with 3p of aluminium).

P22: Because…the 3p orbital is further away from the nucleus…so…the distance…the greater the distance, the attraction is…smaller…energy used to take away the electron…the outermost electron will be smaller.
I: How do you know that the 3p electron is further away?
P22: Using the Aufbau principle.
I: And how does that help you to decide it is further away?
P22: The triangle…you know, where the 1s, 2s, 2p, 3s, 3p, and then arrow…I don't know why…the way you write $1s^2$, $2s^2$, then $2p^3$, $3s^2$, $3p^1$.
I: So you equate that as distance away.
P22: Yes.
I: OK what about you P23?
P23: Same reason as P22…but how I say the distance is different…I assume the distance by seeing the energy diagram.
I: OK…the energy diagram tells you the distance away from the…
P23: Some sort…something like that.

It would seem from the above excerpts of the interviews that students might have the idea that the 3p electron of aluminium was further away than the 3s electron of magnesium from the way they were taught to 'fill' electrons in various orbitals of an atom according to the Aufbau principle, using energy level or 'electrons-in-box' (Hill et al., 1980) diagrams (see Figure 2), or notations such as '$1s^2$ $2s^2$ $2p^6$ $3s^2$ $3p^1$'. The diagrams or notations indicate energy levels, not distance away from the nucleus, so teachers need to be more explicit in their explanations of what the diagrams mean when they use such diagrams. Item 5 had a similar option, which stated that sodium had a higher first ionisation energy than magnesium because the 3s electrons of magnesium were further away from the nucleus (A5). However, it attracted only forty-one students (4%). This could indicate that students took the 3p sub-shell to be further away from the nucleus than the 3s sub-shell; the formalism for showing the pairing up electrons in the same 3s sub-shell of magnesium did not lead to the same

Figure 2. An 'electrons-in-box' or energy level representation of the electronic structure of aluminium



impression of moving away from the nucleus as the representation of entering an arrow into a box in a 'new' sub-shell, the 3p sub-shell of aluminium. Perhaps the same type of inference is drawn by some students when writing the additional '3p' in '…$3s^2$ $3p^1$' for the electronic configuration of aluminium as compared to magnesium, rather than just changing $3s^1$ to $3s^2$ in magnesium compared to sodium.

Another example of students using relation-based reasoning was when they indicated that sodium had a higher first ionisation energy than aluminum because the 3p electron of aluminum experienced greater shielding than the 3s electron of sodium (Item 7, A3, 24%). These students might not have considered the effect of an increase in the nuclear charge of aluminum compared to sodium.

Discussion

Since the students would hardly have encountered the concepts of ionisation energy in everyday life, it was likely that the alternative conceptions arose from the way ionisation energy was taught and learnt (Taber, 2004). The authors believe that the octet rule framework was carried over from the learning of bonding in secondary chemistry (Taber, 1997b, 1999; Tan et al., 1999) – for example, during teaching practice observations, it is common to hear pre-service teachers saying that 'the sodium atom needs to lose an electron to achieve a stable octet electronic configuration', when teaching ionic bonding. As the octet rule framework does not conflict with the accepted scientific concept in explaining why, for example, the second ionisation energy of sodium was higher than its first ionisation energy, students could unsuspectingly hold both the correct concept and alternative conception. They would see nothing wrong with the alternative conception and treat it as an additional explanation for the phenomenon.

Teachers often use 'stable fully-filled or half-filled sub-shell' as a rule-of-thumb to explain the anomaly in the ionisation energy trend across Periods 2 and 3 of the Period Table, and to help students remember the anomaly. A textbook on introductory tertiary chemistry (Lee, 1977) also uses the octet rule framework and stable fully-filled or half-filled sub-shells to explain the anomaly.

> "*The values for Ne and Ar are the highest in their periods because it requires a great deal of energy to break* ***a stable filled shell of electrons****. There are several irregularities. The high values for Be and Mg are attributed to the* ***stability of a filled s shell****. The high values of N and P indicate that* ***a half-filled p level is also particularly stable****. The values for B and Al are lower because removal of one electron leaves* ***a stable filled s shell****, and similarly with O and S* ***a stable half-filled p shell*** *is left*" (Lee, 1977, p. 96, *present authors' emphasis*)

Cann (2000) also commented that this 'half-filled (and also completely-filled) shells having intrinsic stability' reason was common and could be found in textbooks, but it offered "*no explanation in terms of electrostatic or quantum mechanical interactions within the atom*" (p. 1056). As there is no conflict between the curriculum model and the 'stable fully-filled or half-filled sub-shell' reasoning in explaining why the first ionisation energy of magnesium is higher than that of sodium and aluminium, or why the first ionisation energy of phosphorus is higher than that of silicon and sulfur, the 'stable fully-filled or half-filled sub-shell' reasoning is easily accepted by students as an explanation in addition to the curriculum model; it is also easier to remember and quote. Thus, teachers need to be wary of using such heuristics in their teaching.

The conservation of force thinking could have arisen because the students did not integrate their knowledge of electrostatics learned in physics with the concepts of ionisation energy learnt in chemistry (Taber, 1998a, 2003a) or the students might not have studied A-level physics at all. As Taber (2003a) mentioned, conservation of force thinking (like the idea that full shells are desirable) has "*an intuitive attraction to many students*" (p. 156) – one will get a greater portion of a cake if there are fewer people sharing it. It also enables one to predict correctly successive ionisation energies. If only one or two of the three factors influencing ionisation energy (nuclear charge, distance from the nucleus and shielding/ screening effect) were used during lessons to discuss the difference in ionisation energies of two or more elements, then students are not likely to realise that they had to consider all three factors, not just one or two; this could be the cause of relation-based thinking.

It is worth noting that the lack of consistency found in the principles used to answer the earlier questions in the IEDI (Q1-4), and the use of relation-based thinking in the later items (Q5-10), may be closely related phenomena. So those students who, for example, used appropriate electrostatic ideas to answer Q1 but were attracted to the idea that full shells imply stability in Q3, may be selecting what seems the most appropriate response from a repertoire of potentially relevant principles (in itself, a sound strategy, cf. Taber, 1995), in the same way as those who used ideas about, say, increased shielding whilst ignoring increased nuclear charge when comparing elements in Period 3. In one situation the students are selecting from alternatives with different status relative to the curriculum ('alternative' conceptions vs. appropriate concepts), and in the other situation they are selecting only one of the relevant appropriate alternatives – so in both cases they judge one of a number of potentially relevant explanatory principles as 'the' best answer in the context of a particular item.

Teachers need to be aware that they can be the sources of alternative conceptions, for instance, by the way they teach – using the 'stable fully-filled or half-filled sub-shell' heuristics. Teachers can also have the same alternative conceptions as students (Wandersee et al., 1994; Chang, 1999; Lin et al., 2000; Tan, 2005) and can unwittingly pass their own alternative conceptions to their students, or think that there is nothing wrong with their students' alternative conceptions. Pre-service teachers' understanding of ionisation energy is being investigated as an extension of this study; the IEDI has so far been administered to 105 pre-service secondary chemistry teachers, and four pre-service teachers were interviewed using the IEDI as the protocol. When the four pre-service teachers were asked in interviews to explain the trend of ionisation energy across the elements, sodium to aluminium, and silicon to sulfur, and all of them referred to the 'stable fully-filled s sub-shell' and 'stable half-filled p sub-shell' heuristics in addition to the correct concepts in their explanations. Teachers should also realise that textbooks also can contain errors and misleading or conflicting illustrations and statements that can give rise to alternative conceptions (Wandersee et al., 1994; Boo, 1998; de Posada, 1999; Sanger & Greenbowe, 1999).

The results and conclusions generated in this study refer specifically to the sample groups involved in the study. Generalisation of the findings to all A-level chemistry students in Singapore must be considered with caution due to the nature and the limited number of A-level institutions involved in the study. Not all concepts and propositions related to A-level ionisation energy were measured by the IEDI, so the conclusions refer specifically to the concepts and propositions examined by the test items. There are also problems associated with the pencil-and-paper tests (Towns et al., 1993). For example, multiple-choice tests "*make some demands on the reading/comprehension skills of the respondents*" (Taber, 1999, p. 99), and students do not "*always perceive and interpret test statements in the way that test designers intend*" (Hodson, 1993, p. 97). Students may not understand or may misinterpret the questions and options in the IEDI, and since they have little recourse for clarification, this may affect the validity and reliability of the test. However, the interviews included in the present study provide triangulation for the IEDI, and suggest this was not a major problem in the present research.

Conclusions: implications for teaching and research

The findings from this application of a two-tier diagnostic instrument developed and administered to Grade 11 and 12 chemistry students lead us to make a number of recommendations and suggestions relating to the teaching of this topic, and to the direction of further research. It seems clear from this study that many A-level chemistry students in Singapore (as in the UK) have significant difficulties in building up an understanding of ionisation energies that matches the target knowledge in the curriculum. This implies that the current approach to teaching this topic is ineffective. Our research leads to some suggestions for how teachers may adjust their teaching of this topic, but also raises questions about its place in the curriculum that indicate the need for further research.

One key area highlighted in this study was how students apply invalid explanatory principles based on the inherent stability of octets or full shells, and on notions of sharing-out of nuclear attraction. We have explained how the octet rule framework is likely to have been developed during earlier secondary education, and is often encouraged by the language and forms of explanations used by some teachers and textbooks. Teachers need to be careful that when discussing ionisation energy they do not use metaphorical language such as atoms ‘wanting’ complete shells that may imply that these shells have an inherent stability. However, as many students will already be primed to think in this way from previous studies, it is also important for teachers to be very alert and spot when students demonstrate this type of thinking, so they may challenge any statements of this form to create dissatisfaction in the students with their alternative conception in order for conceptual change to take place (Posner et al., 1982). As students readily assign similar stability to full sub-shells or half-filled sub-shells, the same advice applies in these contexts. Anthropomorphic language has its place in science and in learning science (Taber et al., 1996), so teachers should be careful not to ridicule student comments about what atoms ‘want’ or ‘need’. However, teachers should always use such student comments as an opportunity to develop a more scientific understanding: responding by asking the student if they can rephrase their point in more technical language. If not, the teacher should model an explanation using scientifically valid ideas, in the appropriate language.

Teachers should emphasize electrical interactions at all times, and make connections with basic electrostatic (i.e. Coulombic) principles to challenge the common notion that the nucleus gives out a set amount of force, that is somehow shared around the electrons. Teachers should also be careful that whenever using formalisms such as the electrons-in-boxes representations, they should be explicit about the significance of the representation, and check that students appreciate which is signified.

The second area of concern from this study is the way students coordinate (or fail to coordinate) different potential factors that may be significant in making comparisons. What seems clear from the pattern of responses in this study, is that many students seem to be aware of a range of potentially relevant factors (some scientifically valid, some not) that can influence ionisation energy. However, they may be using a faulty ‘search’ strategy (i.e. ‘identify a single relevant factor, and apply it’), or are actually identifying all the relevant factors, but are not able to coordinate them effectively. In the latter case, we need to appreciate whether this is simply a lack of having been taught a suitable strategy - in finding a way to overcome limitations in working memory, (cf. Tsaparlis, 1994, 1998), or whether it indicates a mismatch between cognitive abilities and demands (cf. Shayer et al., 1981). We suspect that when many students are asked to make a comparison between two ionisation processes, they bring to mind one apparently relevant factor and apply it on an ‘all other things being equal’. Thus, they may notice that a magnesium atom has one more electron than a sodium atom but not consider that it also has one more proton in its nucleus.

More research is needed to explore this issue, but for the moment our advice to teachers is that when discussing ionisation energies, and making comparisons, they should always be explicit about all the factors that should be considered (even when some are not relevant for a particular comparison). So, for example, in comparing the second and third ionisations of sodium the teacher should ensure that nuclear or core charge is considered, even though in this case it will be seen that this has not changed and so can then be put aside as a consideration. In this way, the teacher can model the process of always identifying the potential factors at work, and so making decisions about which need to be taken into account in any particular case. Sometimes some factors can be ignored, sometimes influencing factors have effects in the same direction, and sometimes there is a degree of compensation – in this case a student can decide on the dominant factor(s) only by looking at the experimental data.

Ionisation energy is a topic that has historically featured in courses at this level, but this inclusion should be questioned if educational research suggests that most learners are unlikely to cope with the concepts at this stage of their scientific education. It may be that many 16-19 years olds do not yet have the cognitive ability to coordinate a range of factors, in the context of formal models of atomic structure, or (as implied above) perhaps they lack support in developing strategies that would enable them to respond. Either difficulty assumes that students have the required knowledge, *and* can access it for processing. Research into A-level students’ understanding of ‘orbital’ models of the atoms (Taber, 2004) suggests another possibility. Here it was found that even when students could clearly demonstrate they had acquired an ‘orbital’ model of the atom that matched that presented as curriculum knowledge, they had difficulty applying that model in appropriate contexts. It was suggested that although the new knowledge was established well enough to be recalled when cued, it was not yet robust enough to act as the basis of further learning. This conjecture derives from research into memory formation, which indicates that consolidation of new learning typically occurs over a time-scale of many months (see Taber, 2003b). The content of school science is determined by a range of considerations, and is not always ‘educationally sound’ (Kind & Taber, 2005). Educational research should provide the educational community with the basis for making choices about curriculum content, and in designing the curriculum models that are appropriate for different learners.

Acknowledgements

This research was funded by the Academic Research Fund, National Institute of Education, Nanyang Technological University, Singapore, RP 8/00 TKC

References

Ausubel D., (1968), *Educational psychology: a cognitive view*, Holt, Rinehart and Winston, Boston.

Ausubel D.P., (2000), *The acquisition and retention of knowledge: a cognitive view*, Kluwer Academic Publishers, Dordrecht.

Barker V. and Millar R., (2000), Students’ reasoning about basic chemical thermodynamics and chemical bonding: what changes occur during a context-based post-16 chemistry course? *International Journal of Science Education*, **21**, 645-665.

Boo H.K., (1998), Students’ understanding of chemical bonds and the energetics of chemical reactions, *Journal of Research in Science Teaching*, **35**, 569-581.

Cann P., (2000), Ionization energies, parallel spins, and the stability of half-filled shells, *Journal of Chemical Education*, **77**, 1056-1061.

Caravita S. and Halldén O., (1994), Re-framing the problem of conceptual change, *Learning and Instruction*, **4**, 89-111.

Carr M., (1984), Model confusion in chemistry. *Research in Science Education*, **14**, 97-103.

Carr M., (1996), Interviews about instances and interviews about events, in D.F. Treagust, R. Duit, and B.J. Fraser (Eds.), *Improving teaching and learning in science and mathematics*, Teachers College Press, New York, pp. 44-53.

Chang. J.Y., (1999), Teachers college students' conceptions about evaporation, condensation, and boiling, *Science Education*, **83**, 511-526.

De Jong, O. and Treagust, D.F., (2002), The teaching and learning of electrochemistry, in J.K. Gilbert, O. De Jong, R. Justi., D.F. Treagust and K.H. Van Driel (Eds.), *Chemical education: towards research-based practice*, Kluwer Academic Publishers, Dordrecht, pp. 317-337.

de Posada J.M., (1999), The presentation of metallic bonding in high school science textbooks during three decades: science education reforms and substantive changes of tendencies, *Science Education*, **83**, 423-447.

Driver R. and Easley J., (1978), Pupils and paradigms: a review of literature related to concept development in adolescent science students, *Studies in Science Education*, **5**, 61-84.

Driver R., Leach J., Millar R. and Scott P., (1996), *Young people's images of science*, Open University Press, Buckingham.

Driver R. and Oldham V., (1986), A constructivist approach to curriculum development in science, *Studies in Science Education*, **13**, 105-122.

Driver R., Squires A., Rushworth, P. and Wood-Robinson, V., (1994), *Making sense of secondary science: research into children's ideas*, Routledge, London.

Duit R. and Treagust D.F., (1995), Students' conceptions and constructivist teaching approaches, in B. J. Fraser and H. J. Walberg (Eds.), *Improving science education*, The National Society for the Study of Education, Chicago, Illinois, pp. 46-69.

Fetherstonhaugh T. and Treagust D.F., (1992), Students' understanding of light and its properties: teaching to engender conceptual change, *Science Education*, **76**, 653-672.

Garnett P.J., Garnett, P. J., and Hackling, M. W., (1995), Students' alternative conceptions in chemistry: a review of research and implications for teaching and learning, *Studies in Science Education*, **25**, 69-95.

Harrison A.G. and Treagust D.F., (2002), The particulate nature of matter: challenges in understanding the submicroscopic world, in J.K. Gilbert, O. De Jong, R. Justi., D.F. Treagust and K.H. Van Driel (Eds.), *Chemical education: towards research-based practice,* Kluwer Academic Publishers, Dordrecht, pp. 189-212.

Hill G.C. and Holman J.S., (1980), *Chemistry in context*, The English Language Book Society and Nelson, Frome and London.

Hodson D., (1993), Re-thinking old ways: towards a more critical approach to practical work in school science, *Studies in Science Education*, **22**, 85-142.

Kind V. and Taber, K.S, (2005), *Science: teaching school subjects 11-19*, Routledge, London.

Lee J.D., (1977), *A new concise inorganic chemistry*, (3rd Ed.), Van Nostrand Reinhold, Wokingham, Berkshire.

Lin H.S., Cheng H.J. and Lawrenz, F. (2000), The assessment of students' and teachers' understanding of gas laws, *Journal of Chemical Education*, **77**, 235-237.

Novak J.D., (1996), Concept mapping: a tool for improving science teaching and learning, in D.F. Treagust, R. Duit and B.J. Fraser (Eds.), *Improving teaching and learning in science and mathematics*, Teachers College Press, New York, pp. 32-43.

Osborne R.J., Bell B.F. and Gilbert J.K., (1983), Science teaching and children's view of the world, *European Journal of Science Education*, **5**, 1-14.

Palmer D.H., (1999), Exploring the link between students' scientific and nonscientific conceptions, *Science Education*, **83**, 639-653.

Pedrosa M.A. and Dias, M.H., (2000), Chemistry textbook approaches to chemical equilibrium and student alternative conceptions, *Chemistry Education Research and Practice*, **1**, 227-236.

Peterson R.F., (1986), *The development, validation and application of a diagnostic test measuring year 11 and 12 students' understanding of covalent bonding and structure*, Unpublished Master's thesis, Curtin University of Technology, Western Australia.

Peterson R.F., Treagust D.F. and Garnett, P., (1989), Development and application of a diagnostic instrument to evaluate grade –11 and –12 students' concepts of covalent bonding and structure following a course of instruction, *Journal of Research in Science Teaching*, **26**, 301-314.

Posner G.J., Strike K.A., Hewson P. and Gertzog W.A., (1982), Accommodation of a scientific conception: toward a theory of conceptual change, *Science Education*, **66**, 211-227.

Sanger M.J. and Greenbowe T.J., (1999), An analysis of college chemistry textbooks as sources of misconceptions and errors in electrochemistry, *Journal of Chemical Education,* **76**, 853-860.

Schmidt H.J., (2000), Should chemistry lessons be more intellectually challenging? *Chemistry Education Research and Practice*, **1**, 17-26.

Shayer M. and Adey P., (1981), *Towards a science of science teaching: cognitive development and curriculum demand*, Heinemann Educational Books, Oxford.

Taber K.S., (1995), An analogy for discussing progression in learning chemistry, *School Science Review*, **76**, 91-95.

Taber K.S., (1997a), *Understanding chemical bonding – the development of A-level students' understanding of the concepts of chemical bonding*, PhD thesis, University of Surrey.

Taber K.S., (1997b), Student understanding of ionic bonding: molecular versus electrostatic framework? *School Science Review*, **78**, 85-95.

Taber K.S., (1998a), The sharing-out of nuclear attraction: or "I can't think about physics in chemistry", *International Journal of Science Education,* **20**, 1001-1014.

Taber K.S., (1998b), An alternative conceptual framework from chemistry education. *International Journal of Science Education*, **20**, 597-608.

Taber K.S., (1999), Ideas about ionisation energy: a diagnostic instrument, *School Science Review*, **81**, 97-104.

Taber K.S., (2000), Multiple frameworks? Evidence of manifold conceptions in individual cognitive structure, *International Journal of Science Education,* **22**, 399-417.

Taber K.S., (2001), Building the structural conception of chemistry: some considerations from educational research, *Chemistry Education Research and Practice,* **2**, 123-158.

Taber K.S., (2002), A core concept in teaching chemistry, *School Science Review*, **84**, 105 -110.

Taber K.S., (2003a), Understanding ionisation energy: physical, chemical and alternative conceptions, *Chemistry Education Research and Practice,* **4**, 149-169.

Taber K.S. (2003b) Lost without trace or not brought to mind? A case study of remembering and forgetting of college science, *Chemistry Education: Research and Practice,* **4**, 249-277.

Taber K.S., (2004), Learning quanta: barriers to stimulating transitions in student understanding of orbital ideas, *Science Education,* **89**, 94-116.

Taber K.S. and Watts, M., (1996), The secret life of the chemical bond: students' anthropomorphic and animistic references to bonding, *International Journal of Science Education,* **18**, 557-568.

Taber K.S. and Watts M., (2000), Learners' explanations for chemical phenomena, *Chemical Education Research and Practice*, **1**, 329-353.

Tan K.C.D., (2005), Pre-service teachers' conceptions of basic inorganic qualitative analysis, *Canadian Journal of Science, Mathematics and Technology Education*, **5**, 7-20.

Tan K.C.D., Goh N.K., Chia L.S. and Taber K.S., (2003), Ions and ionisation energy, *Australian Journal of Education in Chemistry,* **62**, 21-26.

Tan K.C.D., Goh N.K., Chia, L.S. and Treagust, D.F., (2002), Development and application of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis, *Journal of Research in Science Teaching,* **39**, 283-301.

Tan K.C.D. and Treagust D.F., (1999), Evaluating students' understanding of chemical bonding, *School Science Review*, **81**, 75-83.

Towns M.H. and Robinson, W.R., (1993), Student use of test-wiseness strategies in solving multiple choice chemistry examinations*, Journal of Research in Science Teaching,* **30**, 709-722.

Treagust D.F., (1995), Diagnostic assessment of students' science knowledge, in S.M. Glynn and R. Duit. (Eds.), *Learning science in the schools: research reforming practice,* Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 327-346.

Treagust D.F., Duit R. and Fraser B.J., (1996), Overview: research on students' preinstructional conceptions – the driving force for improving teaching and learning in science and mathematics, in D.F. Treagust, R. Duit and B.J. Fraser (Eds.), *Improving Teaching and Learning in Science and Mathematics*, Teachers College Press, New York, pp. 1–14.

Tsaparlis, G., (1994), Blocking mechanisms in problem solving from the Pascual-Leone's M-space perspective, in H.J. Schmidt (Ed.), *Problem solving and misconceptions in chemistry and physics,* International Council of Association for Science Education, Dortmund, pp. 211-226.

Tsaparlis G., (1998), Dimensional analysis and predictive models in problem solving, *International Journal of Science Education*, **20**, 335-350.

Tyson L., Treagust D.F. and Bucat R.B., (1999), The complexity of teaching and learning chemical equilibrium, *Journal of Chemical Education*, **76**, 554-558.

Voska K.W. and Heikkinen H.W., (2000), Identification and analysis of student conceptions used to solve chemical equilibrium problems, *Journal of Research in Science Teaching*, **37**, 160-176.

Wandersee J.H., Mintzes J.J. and Novak J.D., (1994), Research on alternative conceptions in Science, in D. L. Gabel (Ed.), *Handbook of research on science teaching and learning,* Macmillan, New York, pp. 177-210.

The Appendixes associated with this paper can be found as separate PDF files at http://www.rsc.org/Education/CERP/issues/2005_4/index.asp

Computer aided self assessment – an effective tool

Roy Lowry

Centre for Chemical Sciences, University of Plymouth
Email: R.Lowry@plym.ac.uk

Received 11 July 2005, accepted 26 September 2005

Abstract: Computer aided assessment (CAA) has been used to provide students with a system for formative self-assessment. Students could access the material at any time and two levels of feedback sought to guide further learning. Comparison of two groups of students (those who used all of the material and those who used none) revealed that those students who used the system performed significantly better in the end of module summative assessment. The difference in performance was not observed between these two groups in a similar assessment on material that was not supported by CAA. This points towards the conclusion that the CAA system has made a positive impact upon the learning experience of the students. [*Chem. Educ. Res. Pract.*, 2005, **6** (4), 198-203]

Keywords: effective learning, formative assessment, feedback self-assessment, multiple-choice assessment.

Introduction

Computer Aided Assessment (CAA) is being used increasingly to provide a quick method of marking summative assessments for large groups of students. Whilst this can be very effective in saving time for staff, it does mean that the feedback present in the more traditional coursework assignments is lost. For effective learning, it is important that students can try out their understanding and obtain constructive criticism, so that the learning cycle is complete (Kolb, 1975). There is, therefore, a need to provide feedback to individual students via formative assessments. However, this is an exercise that is very time consuming, especially for large classes. Various authors have attempted to address this using CAA. Whilst this can be purpose-written software for a particular course or module (e.g. Hunt, 2002), software suites for assessment are now available either as part of a managed learning environment (e.g. WebCT, Blackboard) or for assessment alone (e.g. QuestionMark, WebMCQ).

The use of CAA for formative assessment affords considerable advantages. For the institution, this is mainly time saving (after the initial 'cost' of setting up the system), but the advantages for the students are more numerous:

- to give students feedback;
- to guide student effort;
- to diagnose problems in learning;
- to give students experience in assessment methods.

For an excellent discussion of the issues and impacts of using CAA for formative assessment see Charman (1999).

One possible problem with the implementation of a computer-based system (formative or summative) is that the results could be affected by students' prior computing experience or anxiety about using the technology. However, work comparing assessment performance using both computer and paper based multiple-choice tests (Lee, 2001) has demonstrated that there is no measurable effect.

This project sought to provide a formative self-assessment mechanism via CAA for our Stage I BSc Environmental Science students. The syllabus covered included solution chemistry and thermodynamics, and was a 10-lecture 2-workshop course, which was 50% of a 20-credit module in the first semester. These students traditionally shy away from this material as it is seen as ‘hard’ science and contains more mathematics than other disciplines. Thus, any mechanism that increases their interaction with the subject could be beneficial. Also, since this module was at the very beginning of their course, the course team wanted to encourage students to become ‘deep’ rather than ‘surface’ learners (Marton, F., 1976). It was hoped that self-assessment would help to promote this. One hundred and four students completed the module.

Method

As the summative assessment for this module was a multiple-choice test, it was decided that the formative self-assessments should also be in the multiple-choice format. In this way, students could practise both their understanding of the subject and the mode of the summative assessment to come. Using a support mechanism that mimics that of the final assessment should also improve the validity of the assessment as ‘false negatives’ due to lack of familiarity with the method are minimised.

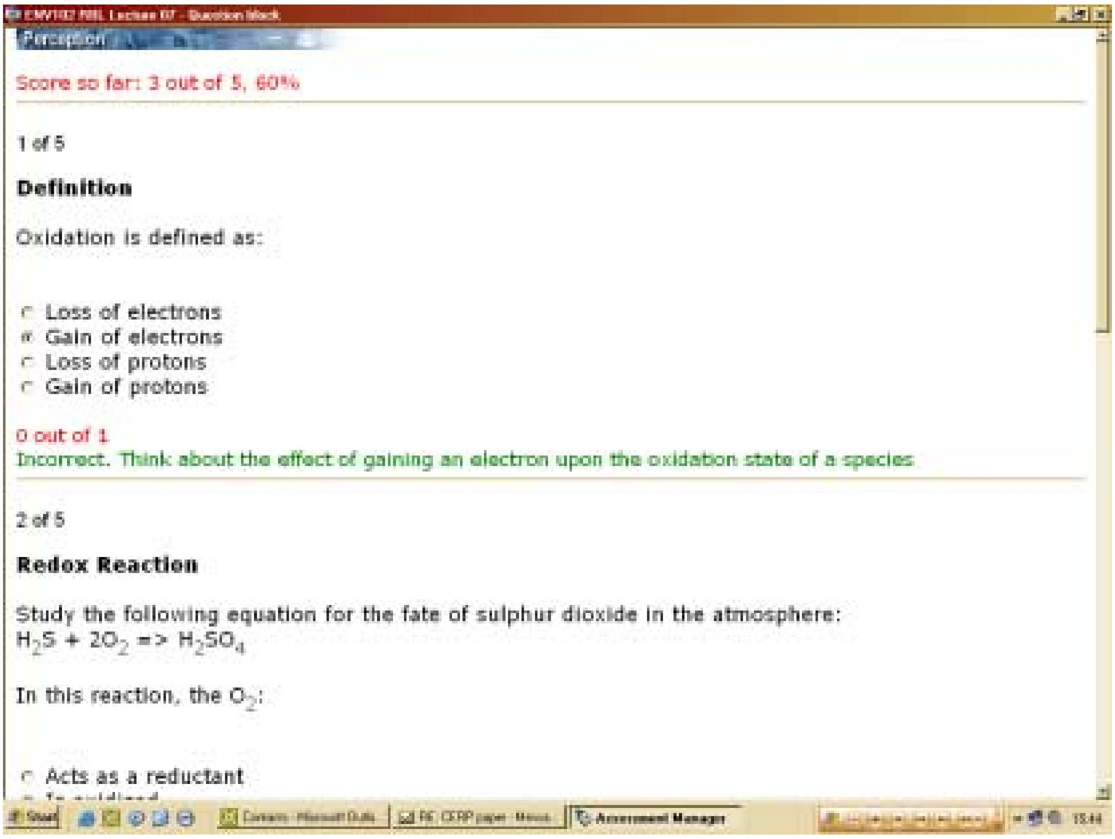
The CAA system used was ‘Perception’ from Question Mark Computing. The system comprised a series of programs for question creation, assessment compilation and delivery/monitoring of the assessments. This last program was mounted on a server (the others are local programs) and supplied the assessments to the students as web pages, allowed different levels of security and collected data as to which students had performed the assessments and their scores (both at assessment and question level). Whilst the suite of programs has been improved over recent years, the implementation of the system at the University of Plymouth is very similar to that reported by Zakrezewski (1999) at the University of Luton.

For each lecture, a short (5 question) self-assessment test was compiled. The test became available via the web at the end of the lecture and could be accessed whenever the student required and as many times as the student wanted. Access was not restricted to just the computers on campus and thus students could use the system from wherever they happened to be – especially during the vacation. At the end of the self-assessment test, the student was given feedback at two levels. The overall feedback gave the score gained and a message based upon the mark. These messages were as follows:

- 100% to 95% “*All correct. Well done! Celebrate your success and then use your valuable study time upon another area*.”
- 94% to 50% “*There are some misunderstandings in your knowledge of this area. Use the individual feedback to identify these before further study. Come back to repeat the test if you wish.*”
- 49% to 0% “*Your understanding of this area requires further work. Go through your lecture notes with a text book and then re-try the test.*”

After this page, students could access feedback relating to their answer to each question. This feedback was constructed so that it explained why an answer was incorrect, but not so that it gave the correct answer. The idea here was to get the students to consider their understanding and not just to memorise. Figure 1 shows a screen shot of one of the self-assessment tests. The end of year summative test consisted of 30 questions to be attempted within 45 minutes and was held under normal examination conditions. It was marked by an optical mark reader.

Figure 1. Screen shot of one of the CAA formative assessments



Score so far: 3 out of 5, 60%

1 of 5

Definition

Oxidation is defined as:

- ○ Loss of electrons
- ● Gain of electrons
- ○ Loss of protons
- ○ Gain of protons

0 out of 1

Incorrect. Think about the effect of gaining an electron upon the oxidation state of a species

2 of 5

Redox Reaction

Study the following equation for the fate of sulphur dioxide in the atmosphere:

$H_2S + 2O_2 => H_2SO_4$

In this reaction, the O_2:

- ○ Acts as a reductant

Results

Although the self-assessment tests were completely formative and non-compulsory, 42% of the students used them all at least once and 65% used at least some of the tests. This high level of use may be due to the similarity between the formative and summative assessment methods (thereby increasing the value to the students). Another possibility is the relatively mature nature of this group of students; 32% of this cohort of students were over 21 years of age when they started the course. It is interesting to note that the majority of those students who accessed all of the tests started using the system within the first two weeks. It would seem that encouraging students to form the habit of testing their understanding early in the course is beneficial as it reduces the possibility of such activities being squeezed out by assessments in other modules. Another factor that may have increased the use of the system was that students were told that the system was not connected to the university record system and that it was ‘safe’ to try out their understanding multiple times without fear that this would affect future grades.

In order to gauge the success of the system, the summative results of two groups of students were compared. The first group (Group 1) had accessed all the self-assessment tests at least once, whilst the second group (Group 2) had not attempted any of the tests. This information was automatically collected by the system and was downloaded into an Excel spreadsheet for analysis. The average result for the end of module summative assessment for these two groups was found to be 52% for Group 1 and 36% for Group 2. Figure 2 shows the frequency histogram of these two groups and clearly there is a difference in performance. To check the significance of this difference, a one-tailed t-test was applied to the data using an Excel spreadsheet, a summary of which is given in Table 1. This confirms the difference to be statistically significant at the 95% confidence level.

Figure 1: Summative assessment results

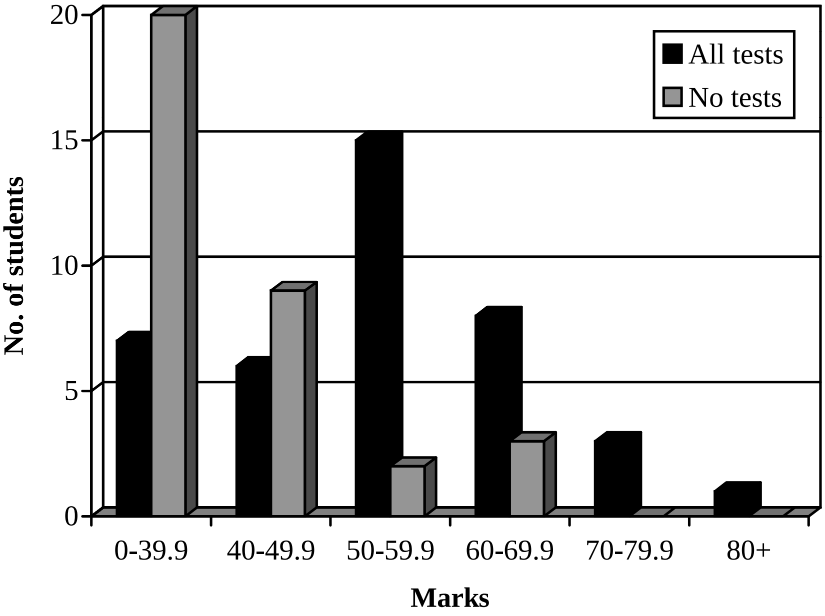


Table 1. Statistical evaluation of results

	CAA supported		Unsupported	
	Group 1	Group 2	Group 1	Group 2
No. of students	40	34	40	34
Average score (%)	52.3	36.1	66.8	63.3
Standard deviation	14.1	13.3	18.6	18.9
t_{stat}	5.001		0.804	
t_{crit} (one tail)	1.667			

Comparison with other material

It is possible that the difference shown between the two groups was due to the nature of the students themselves. Since the groups of students were self-selecting, it could be that the more able or dedicated students elected to use the self-assessment tests, whilst others did not. If this were true, then these students could have performed better than their colleagues without the self-assessment CAA system being in place. To check whether or not that was the case, a similar analysis was performed on the results of another summative assessment, which dealt with material that was similar (inorganic chemistry) but was not supported by the self-assessment CAA system. In common with the previous material, the optical mark reader was used to mark the summative assessment for this series of lectures. The summative marks for this material were analysed using the same groups as previously, i.e. those students that had used all of the available self-assessment tests for the previous material (Group 1) and those that had not used any (Group 2). Using these same groups of students, the means of the two groups differed by only 3.6%. The subsequent t-test confirmed that the difference between these two groups under these circumstances was not significant.

Examining results from these different areas of material results in a comparison between material that has a support mechanism in place (self-assessment via CAA) and material that does not. It could be said that any support mechanism should result in an increase in performance of the student body. The data presented in this paper demonstrate that it is likely that self-assessment CAA is suitable as a support mechanism for material of this type and is an effective tool. However, it should be noted that there were inevitable differences in both the type of material covered and assessed (it is possible that the summative assessment for the supported material was less discriminatory) and also the presentation of the material (two different lecturers). These differences ensured that the unsupported material was not a totally effective 'control'.

A similar study (Peat, 2002) based upon a variety of delivery modes for computer-based self-assessment for biology students also found that formative material provided via computer had a positive influence upon learning. Student feedback over a number of years was consistently positive, but this work did not include a statistical analysis of summative results.

Student perception

Whilst no formal gathering of data on how the students felt about the system was performed, many students who had used the system offered their opinions via various methods, including feedback from the student representatives on the staff-student liaison committee, and the end of module student questionnaire. One of the most common comments was that the system increased confidence. It would seem from this that many students think that they do not understand a concept, when in fact they do. This means that they spend time studying the particular issue when their time could be better spent reading about another area (which they think that they do understand, but possibly do not). In addition, this increase in confidence can mean that students are more open when they arrive in a future lecture, as their perception of the subject as a whole influences their learning (Johnstone, 1997).

Other comments included the usefulness being able to access the material at any time and not having to come into the university. These factors may well have contributed to the considerable uptake of the self-assessments.

Conclusion

The data collected in this study indicate that providing a CAA system for self-assessment positively affects the learning of those students who choose to use it. Once the question database has been set up, no further intervention is required by staff, apart from scheduling the assessments. Given this, the time required to set up the system is more than justified by the improvement in the learning of the students. However, this study does not contain a true control group and hence it is possible that the effect seen upon summative assessment performance was due to other factors.

Further work

The system has been successful in providing formative assessment and therefore we will move to using it for the summative assessment in the coming academic year. This will allow a much greater range of question types to be used in both the formative and summative modes by removing the limitations imposed by the use of the optical mark reader for the summative assessments. In particular, multiple response (more than one correct answer from a variety of choices), ranking (place items in the correct order) and numeric questions are suitable for material of this type. For a discussion of possible question types and their relative merits/demerits see Clarke, 2001. The system will also be extended to cover all of the material in the module and the results of the study reported here used to encourage more students to use the system.

References

Charman D., (1999), Issues and impacts of using computer-based assessments (CBAs) for formative assessment in Brown, S., Race, P. and Bull, J. (eds) *Computer-assisted assessment in higher education* London: Kogan Page.

Clarke A., (2001), *Designing computer-based learning materials*, Gower, Aldershot, 61-78.

Hunt N., Hughes J. and Rowe, G., (2002), Formative automated computer testing, *British Journal of Educational Technology,* **33**, 525-535.

Johnstone A.H., (1997), …And some fell on good ground, *University Chemistry Education*, **1**, 8-13.

Kolb D.A. and Fry R., (1975), Toward an applied theory of experiential learning, in C. Cooper (ed.) *Theories of group process*, John Wiley, London.

Lee G. and Weerakoon P., (2001), The role of computer-aided assessment in health professional education: a comparison of student performance in computer-based and paper-and-pen tests, *Medical Teacher,* **23**, 152-157.

Marton F. and Saljo R., (1976), On qualitative differences in learning I. Outcome and process, *British Journal of Educational Psychology,* **46**, 4-11.

Question Mark Corporation Home Page http://www.questionmark.com/uk/home.htm (accessed September 2005).

Zakrezewski S. and Bull J., (1999), The mass implementation and evaluation of computer-based assessments, *Assessment and Evaluation in Higher Education*, **23**, 141-152.

Educational research

Assessment formats: do they make a difference?

Eleni Danili and Norman Reid

Centre for Science Education, University of Glasgow, Glasgow, G12 8QQ, UK
e-mail: N.Reid@mis.gla.ac.uk

Received 23 August 2005, accepted 13 October 2005

Abstract: This study has explored the relationships between the results of various formats of paper-and-pencil classroom assessments in five classroom chemistry tests. The formats of assessment that have been used were: multiple choice, short answer, and structural communication grid. The study was conducted in Greece with the participation of first year upper secondary public school pupils (Lykeio, Grade 10, age 15-16). The correlations between the different formats of assessment tended to be between 0.30 and 0.71. This is a wide range but even the highest value is well short of 1.0. This suggests that the best student found by one method is not necessarily the best student by another method. This raises questions about the validity of the formats of the assessment and what different formats of assessment are testing. [*Chem. Educ. Res. Pract.*, 2005, **6** (4), 204-212]

Key Words: Assessment formats, correlation, paper-and-pencil classroom chemistry assessment.

Introduction

Assessments play an important role in the teaching and learning process, and for specific uses. For individuals, assessments, particularly public examinations, profoundly affect life chances, not just in the first years after leaving school, but many years later. As Boud (1995. p. 35) stated "*the effects of bad practice are far more potent than they are for any aspect of teaching. Students can, with difficulty, escape from the effects of poor teaching, they cannot (by definition, if they want to graduate) escape the effects of poor assessment. Assessment acts as a mechanism to control students that is far more pervasive and insidious than most staff would be prepared to acknowledge*".

Indeed, to evaluate someone and make decision for his/her career and future is not an easy task to do. It is a very difficult one and carries with it awesome responsibility. Therefore, some authors characterizations for assessment were: "*both time consuming and potentially dangerous*" (Johnstone, undated, p. 2); "*a serious and often tragic enterprise*" (Ramsden, 2003, p. 13); "*nightmares*" (Race, 1995, p. 61).

Assessment can take many forms, and ideally, performance should be unrelated to the mode in which a test is administered. However, can this statement be true in a real situation? Evidence from research shows the effects of assessment task format on student achievement (e.g. Caygill and Eley, 2001). Moreover, Friel and Johnstone (1978a) showed 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. Ideally, if a test is reliable (in a test, re-test sense) and same knowledge and understanding is being assessed, the two orders of merit should be actually very similar 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. In that case, if rank-order correlation is worked out between the two orders of merit, this should be 1.0, a perfect match in order. A complete reversal of the order would give a value of -1.0 and a completely random pair of orders would give values tending to zero (Johnstone and Ambusaidi, 2000). Their research found that the rank-order correlation was about 0.6. This suggests that the two rank-orders of merit have in common only about 36% of the variance [$(0.6)^2 = 0.36$ that is 36%]. This suggests that the two orders of merit have some similarity, yet are by no means well matched (Johnstone and Ambusaidi, 2000).

In classroom assessment, the study of Yuh-Yin and I-Fen (2000) with science tests found that the correlation between multiple-choice items and short-answer question was 0.68 in one topic (solutions) and 0.77 in another topic (momentum), while the correlations between the same formats of assessment but in different content areas were smaller. Thus, for multiple-choice solution and multiple-choice momentum, the correlation was found to be 0.47, and for short-answer solution and short-answer momentum it was found 0.66. Moreover, correlations between multiple-choice items and short-answers questions with performance-based assessment were smaller even in the same area of content (0.48 and 0.46 respectively). It was assumed that, within the same content area, multiple-choice and short-answer tests measured similar cognitive components, while performance-based assessment emphasized different cognitive dimensions.

It is certain that no one method of assessment is adequate for testing a course. A battery of test methods is required to allow for a fair measure of our students' attainments (Balla and Boyle, 1994) and "*to cater for the range of student abilities, of testable objectives and student maturity*" (Johnstone, 2003). Indeed, Race (2003) argued that, "*the greater the diversity in the methods of assessment, the fairer assessment is to students*". However, in recent years there has been a temptation to adopt objective testing to cope with the rise in student population (Johnstone, 2003). This situation creates problems, because "*to conduct all assessment by this method is not advisable. The most intellectually mature students generally hate objective testing because they need room to expand and show their independence of thought*" (Johnstone, 2003). Moreover, each one of the formats of assessment can be claimed to disadvantage those students who do not give of their best in the particular circumstances in which it is used. Therefore, diversifying assessment so that students experience a range of assessment methods balances out the situation, and increases the chance that they will be able to demonstrate their best performance in at least some of the formats.

Design of the project

The aim of this study was to examine the correlation between the results of different paper-and-pencil formats of classroom assessment. The aim was to build on previous work by extending the range of formats and using a set of assessment across the work of many months of school study.

The work was conducted in Greece with the participation of first year upper secondary public school pupils (Lykeio, Grade 10, age 15-16) during March-April 2002 and during the school year September 2002 to May 2003. Five chemistry tests were designed and each chemistry test assessed pupils by a range of question formats asking about the same knowledge and understanding in the same topic.

There were difficulties and restrictions in relations to the format of questions that the researcher wanted to apply. For example, the researcher wanted to try not only Structural Communication Grid questions which allow for pattern seeking but also Structural Communication Grid questions which look for sequencing and even for a kind of 'objective essay' (Johnstone, 2003). However, the teachers objected to these questions because they thought pupils were not familiar with them and this might have caused problems.

The most common pencil-and-paper formats of assessment used in educational practice in Greece are:

- Open-ended questions (OE)
- Close-response question or objective tests such as:
    - Multiple-choice Questions (MC)
    - True-false,
    - Matching questions,
    - Identifying reasons to support assertions
    - Filling in blanks to complete statements
    - Grid questions (SCG)
- Short answer question (SA)
- Solving problems (mainly of algorithmic type)

The chemistry topics that were tested followed the timetable and the syllabus of the Greek schools. The tests were constructed after looking at the study questions with the Greek Chemistry textbook (Liodakis, 1999) the Standard Grade Chemistry book (Renfrew, 1995) and the textbook by Moore et al. (1998) in order to develop questions in formats and styles appropriate for the pupils. Thus, the tests were based on:

Test 1:	Atomic structure, classification of matter, solubility
Test 2:	The periodic table and chemical bonds
Test 3:	Mole concept
Test 4:	Acids, alkalis, pH, neutralisation
Test 5:	Solutions

The researcher contacted several teachers of different schools and explained to them the purpose of the project. Schools were selected on the basis of teachers being willing to assist in what was a large project. Table 1 shows the range of question formats that have been used in each test of the project, the number of schools as well as the number of pupils who have been involved in the project.

Table 1: Type of formats of assessment used in each chemistry test and number of pupils have been involved

Chemistry Test	Type of formats	Number of schools	Number of pupils
Test 1	MC-SA	8	288
Test 2	SA-SCG	4	185
Test 3	SA-SCG	3	146
Test 4	MC-SCG-SA	7	321
Test 5	MC-SCG-SA	2	64

MC:	Multiple-choice
SA:	Short-answer (open-ended)
SCG:	Structural communication grid

Only 64 pupils sat the Test 5. One of the reasons for that was that the test was given to the teachers towards the end of the school year and, at that time, the pupils usually are very busy with other activities. Thus, many hours of teaching are lost and teachers are mainly concerned to finish the teaching units and they are not willing to spend time to evaluate and assess the results of their teaching.

Weighting of marks was carefully decided to reflect the demand level of questions. For every section of each test, raw scores were converted into a percentage and these were combined to give the total mark for each pupil. Converting the raw scores to a common scale makes it easier for the reader to compare mean scores between different tests and see patterns that might emerge from the study, though it does not alter the statistical results.

All the test papers were marked by the class teachers who were familiar with both the course and what was a reasonable standard. The test papers were then re-marked by the first author to ensure that standards were maintained. In fact, because there were strict marking schemes, marks were rarely modified on re-marking. It is easy to show that the inter-marker reliability will be very high given a tight marking scheme (typically over 0.95).

The tests were set to match the styles and standards typical in Greece and were based on the actual tests normally used. Mean marks did vary considerably, reflecting what normally happens in schools in Greece where all pupils have to take chemistry and a sizeable minority do *not* wish to take chemistry. In this way, the study reflected closely the reality of what actually happens. There is no reason to suppose that tests of different difficulty will lead to unreliability on its own in that the spread of marks in all tests suggests a good discrimination. Discrimination indices are quoted for the multiple-choice questions in the Appendix.

The aim was to make the various formats of the test in any content area as similar as possible. This is not easy and, in reality, tests seek to test *samples* of work in a content area. Nonetheless, the attempt was made as strictly as was possible.

Description of each chemistry test and statistical results

For each chemistry test, descriptive statistics and correlations between the different formats of questions were calculated. Both Pearson coefficient and Spearman's rho correlation between the formats of questions were calculated and were found to give similar values. However, because the distributions were frequently observed to deviate from the normal distribution, it was decided that the Spearman's rho coefficient was more appropriate and this is used in all subsequent discussion. The tests are shown in the Appendix.

Test 1: Multiple-choice vs. short-answer format

Test 1 was based on the introductory chapter of the Greek chemistry textbook. The content areas that it tested were atomic structure, classification of matter, and solubility. It consisted of two sections:

Section 1: 12 Multiple-Choice questions	12 marks,
Section 2: 5 Short-Answered questions	14 marks.

Section 1 had multiple-choice questions, which mainly require students to recognize or identify knowledge. In section 2 there were short answer questions covering the same thematic area. However, the demands on students were more than simply recognition and memorisation. The short-answer questions varied considerably. For example some required students to recall and define knowledge, others required to solve a numerical problem, which requires only a small numbers of steps but deep understanding of the concept involved, or to interpret a graph. Table 2 shows the descriptive statistics for test 1 and the Spearman's rho correlation between MC section and SA section. As can be seen from the table, the SA test was more difficult than the MC test. The Spearman's rho correlation between the MC and SA scores was found to be 0.71 (significant at the 0.01 level - 1-tailed). This correlation is the highest found in the whole study.

Table 2: Descriptive statistics of Test 1

Test 1	N	Minim.	Maxim.	Mean	S.D.
MC	288	17	100	64.3	20.4
SA	288	0	100	53.5	25.6

Spearman's rho between MC and SA = 0.71
significant at 0.01 level (1-tailed)

Test 2: Short-answer vs. structure communication grid

Test 2 was based on the periodic table and bonding theory chapters of the Greek chemistry textbook. It included two sections.

Section 1: 3 Short-answer questions	10 marks
Section 2: 1 Structural communication grid question	10 marks

In order to answer the test, pupils were allowed to have the periodic table in front of them. Thus, in both sections the questions require no recalling of the scientific facts. All questions require an understanding of taught concepts (the periodic table, the properties of the element and the concept of bonding theory), and an ability to interpret the presented information and to apply it. However, short-answer questions require pupils to use their language skills more, since the questions ask pupils to give explanations, e.g. for properties of compounds or for similarities of elements. Table 3 shows the descriptive statistics for the test 2. As can be seen for the table the SCG test was more difficult than the SA test and the Spearman's rho correlation between SA section and SCG section was found to be 0.38, which is relatively low.

Table 3: Statistics of the Test 2

Test 2	N	Minim.	Maxim.	Mean	S.D.
SA	185	0	100	52.2	30.7
SCG	185	0	100	36.7	25.4

Spearman's rho correlation between SA and SCG = 0.38
significant at the 0.01 level (1-tailed)

Test 3: Short-answer vs. structure communication grid

Test 3 was based on the mole concept and Avogadro's Law. It was a short test and included two sections.

Section 1: 2 Short-answer questions	10 marks
Section 2: 1 Structural communication grid question	10 marks

The questions require retrieval of declarative knowledge and procedural knowledge, as well as numerical problem solving ability (of the algorithmic type) in both formats of assessment

Table 4: Statistics of the Test 3

Test 3	N	Minim.	Maxim.	Mean	S.D.
SA	146	0	100	60.9	37.3
SCG	146	0	100	67.7	36.7

Spearman's rho correlation between SA and SCG = 0.55
significant at the 0.01 level (1-tailed)

Test 4 :Multiple-choice vs. grid vs. short-answer

Test 4 was based on the content area of acids; bases; oxides and neutralisation reactions. It had three sections:

Section 1: 13 Multiple-choice (MC) questions	13 marks.
Section 2: 2 Structural communication grid (SCG) questions	12 marks.
Section 3: 3 Short-answer (SA) questions	14 marks.

Mainly, the questions asked students to recall, define, recognise and apply knowledge.

Table 5: Statistics of Test 4

Test Format	N	Minim.	Maxim.	Mean	S.D.
MC	321	8	100	52.9	19.8
SCG	321	0	100	35.2	23.3
SA	321	0	100	34.5	31.4

Spearman's rho correlations between:
MC and SCG = 0.64
MC and SA = 0.64
SA and SCG = 0.66
all significant at the 0.01 level (1-tailed)

Test 5: Multiple-choice vs. grid vs. short-answer

Test 5 was based on the content area of solutions. It had three sections:

Section 1: 5 Multiple-choice questions	5 marks
Section 2: 1 Structural communication grid question	5 marks
Section 3: 3 Short-answer questions	5 marks

The questions were developed mainly from the chemistry book of Moore et al. (1999), in which the assessment questions test understanding and applying chemical concepts. Thus, the answers to the test did not require much memorisation and recall of chemical concepts but the ability to interpret the given information, and understanding of the concept of concentration in solutions, and how it changes when water is added to the solution or water is evaporated from the solution. It required arithmetic skills for answering the open-ended questions.

Table 6: Statistics of Test 5

Test 5	N	Minim.	Maxim.	Mean	S. D.
MC	64	0	100	67.5	28.6
SCG	64	0	100	68.6	26.4
SA	64	0	100	50.8	35.6

Spearman's rho correlation between:
MC and SCG = 0.46
MC and SA = 0.49
SA and SCG = 0.30
all significant at the 0.01 level (1-tailed)

It was expected that very high correlations between different formats of assessment in this test would be found because it was testing the same narrow area of understanding (concentration of a solution, and how the concentration changes by mixing two solutions or diluting a solution). Thus, it is surprising that the correlations were fairly low.

Comparison between Test 1 and 2 for the same group of pupils

Some pupils sat two tests (1, 2). This gave an opportunity to explore the correlations across content areas for the same and for different formats of questions for this group of pupils. Table 7 shows a correlation matrix between the different formats of questions in each test and for the three formats.

Table 7: Correlations across content areas in Test 1, 2

	Test 1		Test 2	
	MC	**SA**	**SA**	**SCG**
MC Test 1	1.00	0.54**	0.38**	0.26*
SA Test 1	0.54**	1.00	0.58**	0.42**
SA Test 2	0.38**	0.58**	1.00	0.54**
SCG Test 2	0.26*	0.42**	0.54**	1.00

** Correlation is significant at the 0.01 level.

Table 7 shows that the correlations between the different formats of assessment in the *same* content area both have a value of 0.54 (short answer – multiple choice; short answer structural communication grid).

The correlation between the short answer formats of assessment in *different* content areas is 0.58, this latter figure suggesting a reasonable reliability of testing.

Multiple choice in test 1 correlated with structural communication grid in test 2, with a value of 0.26 while short answers in test 1 correlated with structural communication grid in test 2, with a value of 0.42.

In one study a correlation of 0.6 was found between multiple choice and short answer question in the same content area. When the multiple-choice questions were re-marked using partial credit marking, the correlation rose to 0.9 (Johnstone and Ambusaidi, 2001). This suggests that test reliability itself is not the problem but the method of marking may be.

There are several factors that might explain the low correlations:

(a) Test reliability – probably not a major factor;
(b) Content area – pupils may perform differently with different material;
(c) Test structure – in table 7 three structures are used.

In multiple-choice tests, pupils tend to eliminate two responses and make a decision between the remaining two, the test being a test of recognition. In structural communication grids, pupils look at each box in turn and decide whether it contains a possible answer, this involves recognition, although thought may be required (in that the number of possible answers is not known). In short answers pupils have to interpret the question and generate answers by recall or thought. In essence, the three types of test are probing different skills and this probably explains the lack of perfect correlation. However, it does raise issues about what skills are being measured – psychological or knowledge of chemistry.

Conclusions

From all the tests it is clear that pupils' performances in MC section were higher than SCG items and SA items. In addition, in this study, the correlation values ranged from 0.30 to 0.71. The higher values tended to occur when *different formats* tested the *same content area* but none of these approached 1.0. We expected higher correlations between MC and SCG questions because both are objective tests. However, the correlations are fairly low in these comparisons. Thus, differences in correlation did not simply arise because some questions require writing while others are objective (in the sense that only a number or letter is required or a box has to be ticked).

These findings are consistent with those of Friel and Johnstone (1978a), Yuh-Yin (2000) and Badger (1990). This suggests that the best student found by one method is not necessarily the best student by another method. If the two formats of assessment were simply testing the same content, then very high correlation would be expected. This also raises questions about the validity of the formats of the assessment. The main question is, what are the different formats testing? Are the different formats testing different abilities and skills, which involve different cognitive factors? Are the different formats testing chemistry or cognition? Are the ways in which the questions are presented having an impact of the pupils' performance (e.g. the use of pictures, or diagrams)? Thus the fundamental issues arising from the study are:

1. Are the different formats of questions testing different abilities or just different themes in a discipline? Probably both?
2. Is any particular format of assessment more valid than others?
3. Are the different formats related to differences between students in one or more psychological traits?
4. It might be reasonable to suppose that the use of multiple formats of assessment tests students more fairly than the use of a single format but on what basis can this be justified?

Clearly, assessment formats do make a difference to student performance. These results led to another study that sought to explore some of the psychological factors that might account for these difference. The findings of that study will be discussed in a future paper.

References

Balla J. and Boyle P., (1994), Assessment of student performance: a framework of improving practice, *Assessment and Evaluation in Higher Education,* **19**, 17-28.

Boud D., (1995), Assessment and learning: contradictory or complementary? In P. Knight (Ed.), *Assessment for learning in higher education,* London: Kogan Page Ltd.

Caygill R. and Eley L., (2001), *Evidence about the effects of assessment task format on student achievement*, Annual Conference of the British Educational Research Association.

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

Johnstone A.H., (2003), LTSN Physical sciences practice guide: effective practice in objective assessment, Hull, LTSN.

Johnstone A.H., (undated), Unpublished lectures, unpublished manuscript, Centre for Science Education, University of Glasgow.

Johnstone, A.H. and Ambusaidi, A., (2000), Fixed Response: what are we testing? *Chemistry Education Research and Practice,* **1**, 323-328.

Liodakis S., Gakis D., Theodoropoulos D., Theodoropoulos P. and Kallis A., (1999), Chemistry A' Lyceum Athens: Organismos Ekdoseon Didaktikon Biblion.

Moore J., Stanitski C., Wood J., Kotz J. and Joesten, M., (1998), *The chemical world concepts and applications* (2nd ed.), New York, Saunder College.

Race P., (1995), What has assessment done for us and to us? In P. Knight (Ed.), *Assessment for learning in higher education*, London, Kogan Page Ltd.

Race P., (2003), *Designing assessment to improve Physical Sciences learning-exams*, http://www.physsci.ltsn.ac.uk/Publications/PracticeGuide/guide4.pdf [2003, 6/10/2003].

Ramsden P., (2003), *Learning to teach in higher education* (2nd ed.), London, Routledge Falmer.

Renfrew R. and Conquest N., (1995), *Standard Grade Chemistry,* London, Hodder and Stoughton.

Yuh-Yin W. and, I-Fen G., (2000), *Classroom assessment forms and their relations with cognitive components. An example from Taiwan*, Paper presented at the Annual meeting of American Educational Research Association, New Orleans, Louisiana.

The Appendixes associated with this paper can be found as separate PDF files at http://www.rsc.org/Education/CERP/issues/2005_4/index.asp