Vol. 5 No. 1 April 2001 Pages 1-41 ISSN 1369-5614

UNIVERSITY CHEMISTRY EDUCATION

RS•C
ROYAL SOCIETY OF CHEMISTRY

The higher education chemistry journal
of
The Royal Society of Chemistry

University Chemistry Education

The Higher Education chemistry journal of the Royal Society of Chemistry

2001 Volume 5, Issue No 1
ISSN 1369-5614

Contents

Editorial Policy for University Chemistry Education (UChemEd)

The journal is aimed at those who teach chemistry in higher education. As a journal for all practising teachers of chemistry at this level, it deals with any topic of practical relevance and use to those involved. It is a place to publish effective methods and ideas for the teaching and learning of chemistry and issues related to the effectiveness of teaching and learning. Contributions are particularly welcome if the subject matter can be applied widely and is concerned with encouraging active and independent learning, with increasing student motivation for learning, with helping them to become effective exploiters of their chemical knowledge and understanding, or with assessment. Contributions should be of clear practical interest to those who teach chemistry.

There are not hard and fast rules for subdividing manuscripts. However, an introduction should provide a clear statement of the relationship of what is described to previous work and opinion (and is likely to include some references to some aspects of educational theory), and also the overall purpose of the article (including, where appropriate, the educational objectives, intended learning outcomes and why these are not satisfactorily achieved by other approaches). Other sections may be equivalent to *methods*, *results*, and *discussion* as used in conventional scientific papers; these sections would describe how the work was carried out, show or illustrate the outcomes (new teaching material etc) which have been created, and critically evaluate how far the original objectives have been met.

Four types of contributions may be submitted:

Full Papers describe a specific method of or approach to teaching, or some teaching material that has been used by the author; papers should explain the educational objectives that led to the use of the method.

Communications are brief accounts of work still undergoing evaluation and development, but of sufficient interest to merit publication.

Reviews provide for practitioners an up-to-date survey of current methods or approaches to teaching and learning and also show how these relate to our understanding of student learning.

Perspectives provide an opportunity for contributors to present a concise but in-depth analysis of a topic of general interest, with clear conclusions likely to be directly useful to other academics involved in teaching.

Letters: these are a medium for the expression of well-argued views or opinions on any matter falling within the remit of Journal.

All contributions, whether or not they were solicited, are rigorously reviewed. Referees are required to evaluate the quality of the arguments presented, and not to make subjective judgements involving their personal views of what constitutes good or effective teaching. Contributions are judged on:

(i) originality and quality of content;
(ii) the appropriateness of the length to the subject matter;
(iii) accessibility of supporting material

University Chemistry Education

Guidelines for Authors

Submission of contributions

University Chemistry Education (UChemEd) is aimed at those who teach chemistry in higher education. As a journal for all practising teachers of chemistry at this level, it deals with any topic of practical relevance and use to those involved. It is a place to publish effective methods and ideas for the teaching and learning of chemistry and issues related to the effectiveness of teaching and learning.

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, or on a floppy disk, to Dr Stephen Breuer at s.breuer@lancaster.ac.uk or at School of Physics and Chemistry, Lancaster University, Lancaster, LA1 4YB, United Kingdom, or to Professor Patrick Bailey at p.d.bailey@hw.ac.uk or at Department of Chemistry, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS, United Kingdom.

2. Submitted contributions are expected to fall into one of several categories. 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. These are the following.
   Full papers describe a specific method of or approach to teaching, or some teaching material that has been used by the author.
   Communications are brief accounts of work still undergoing evaluation and development, but of sufficient interest to merit publication.
   Reviews provide for practitioners an up-to-date survey of current methods or approaches to teaching and learning and also show how these relate to our understanding of student learning.
   Perspectives provide an opportunity for contributors to present a concise but in-depth analysis of a topic of general interest, with clear conclusions likely to be directly useful to other academics involved in teaching.

   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" 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 as superscripts, designated as numbered endnotes.

   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.

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. Headings should be no more than 40 characters in length and subheadings and sub-sub-headings no more than 30.

   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 general practice in the various titles within the Journal of the Chemical Society range. For example:

   Books and Special Publications:
   - W.G. Perry, *Forms of intellectual and ethical development in college years: a scheme*; Holt, Rinehart and Winston, New York, 1979. p. 218.
   - M. McCloskey in D. Gentner and A.L. Stevens (eds), *Mental models*; Lawrence Erlbaum, New Jersey, 1983, pp. 147-153.

Journal Articles:

- D.C. Finster; *J.Chem.Ed*., 1989, **66**, 659.
- A.H. Johnstone and K.M. Letton, *Educ.Chem*., 1990, **27**, 9.

References to articles in *U.Chem.Ed*. on the Web should be identified by paper number until they are incorporated into numbered issues.

7. All contributions submitted will be refereed anonymously. The decision of the Editors on the acceptance of articles is final.

8. Authors grant *U.Chem.Ed*. 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 as soon as the editorial process is complete. In addition, they will be combined into electronic issues as a PDF file in April and October.

10. **Letters to U.Chem.Ed**. relating to published articles or on any other educational topic should be sent for publication on the U.Chem.Ed. Web Board to its moderator, Dr Paul Yates at p.c.yates@chem.keele.ac.uk. A selection of these letters will be incorporated into the issues of the journal.

Generating Coursework Feedback for Large Groups of Students Using MS Excel and MS Word

Paper

Philip Denton

School of Pharmacy and Chemistry, Liverpool John Moores University, Liverpool, L3 3AF, UK.
E-mail: p.denton@livjm.ac.uk

A novel electronic procedure for generating and returning coursework feedback to students has been introduced by tutors at Liverpool John Moores University. The technique uses a combination of Microsoft Excel 97 and Microsoft Word 97 to generate personalised feedback sheets that can include the student's mark, position in the class, and a series of statements selected from a bank of comments, written by the tutor. Feedback sheets can be printed off and returned to students with their marked work, or distributed via e-mail. This procedure is particularly suited to classes undertaking the same coursework assignment, a common feature of undergraduate chemistry courses, and can make the assessment of work from large groups considerably less onerous. The operation of the software is described and the responses of staff and students to the procedure are reported.

Introduction

The importance of assessment in learning is well documented.[1, 2] It is generally accepted that if students are to gain the maximum educational benefit from a written coursework submission, their marked script should be returned with appropriate annotations. In particular, tutors should indicate to students where they have done well, where their misunderstandings are, and what follow-up work might be required.[3] Such written comments do more to motivate students than ticks or crosses alone. Indeed, Ramsden[4] suggests that "*…beneficial information about progress is valued even more by students than qualities such as clear explanations and the stimulation of interest.*" Accordingly, studies in this area indicate that an absence of feedback is an important contributory cause of student failure.[4]

Although educationally sound, the extensive annotation of students' work requires a considerable investment of time and effort by the assessor. It is understood, however, that marked work should be returned as quickly as possible if students are to pay attention to the marker's comments. Thus, Gibbs and Habesha w state that a few weeks after a coursework submission, students have moved onto another topic and, "*have neither the time or the interest to take feedback to heart.*"[3] The introduction of electronic methods can decrease the time taken for feedback to be returned to students. For example, the use of multiple choice question sheets, where student responses are analysed by an optical mark reader, [5] enable work to be graded rapidly. Such approaches have been criticised, however, if they give the student no way of knowing why they got particular question incorrect.[6] Advanced software packages that require students to answer a series of questions may provide in-depth explanations of answers and direct the student to further reading.[7, 8] It is evident that computer assessments that provide immediate feedback can have a positive effect on student attainment.[9]

The Examine software developed at the University of Nottingham neatly illustrates a drawback of all the computer-assisted methods of assessment that are currently available.[10] Although this package will accept multiple-choice answers, numerical responses and written text passages up to 150 words in length, the latter cannot be marked by computer. This is a major limitation, given that a large part of student assessment in chemistry relies on the grading of written work, such as laboratory reports.

Electronic methods can be employed to generate written feedback to students on work that is assessed by tutors. It is suggested, for example, that a word processor is used to build up a bank of feedback comments, which can be copied and pasted into personalised feedback sheets along with general comments relating to the class performance.[11] Presumably, however, this would require the tutor to undertake a number of tedious cut and paste operations. Ideally one would want a system that could automatically generate large amounts of individualised feedback, after tutors had entered the minimum possible amount of information relating to the assessed exercise.

At present, there appears to be a dearth of commercially available software for generating written comments to university students on their assessed work. The development of *electronic feedback*, a software package based on MS Excel 97 and MS Word 97, is a response to this recognised need. The programme has an additional advantage in that it allows tutors to distribute feedback via e-mail. The purpose of this paper is to describe this method and to report on its effectiveness when marking chemistry coursework at Liverpool John Moores University (JMU). Given that Excel and Word are applications that have a wide user base and that most institutions have well developed e-mail networks, it is thought that this procedure will be readily transferable.

Method

The electronic feedback software consists of two programs: *Feebac5.xls*, an MS Excel 97 workbook, and *Fb.doc*, a MS Word 97 document. Together, these programmes can be used by tutors to generate individual word-processed feedback reports that can be printed and/or e-mailed to each member of the class. Each feedback sheet details the student's name and can include the percentage mark, class rank, a general comment, and a series of comments directed specifically at the student. To illustrate the operation of the software, a fictitious set of data has been created. This data set is smaller in size than one that might typically be considered, but is sufficient to convey the essential details of the procedure. An example feedback sheet that has been generated using this data set is shown in Figure 1.

Preparation

The Excel workbook *Feedbac.xls* is composed of a series of worksheets; *Configure, List, Header, Annos*, *Numbers* and *Report*, that contain a number of blank cells. The feedback sheets are created using data that is entered into these cells by the assessor. Tutors enter student names, e-mail addresses and registration numbers into the *List* worksheet, Figure 2. Typically, this information is readily transferred from institutional electronic information services. The class list shown in Figure 2 has the format preferred by JMU in which data appears in the order; e-mail address, forename, surname and registration number. Users can configure the software so that it will accept

FEEDBACK SHEET Created at 13:23 pm on 18/9/2000

Determination of a rate constant for the reaction of I^-(aq) and $S_2O_8^{2-}$(aq).

Assessed by Dr. Philip Denton

STUDENT:	**CATHERINE BAKER**
MARK:	**24 % (HIGHEST: 76 %, AVERAGE: 46 %, LOWEST: 24 %)**
RANK:	**8th out of 8**
COMMENT:	**Satisfactory work. This work was submitted late. A lateness penalty has been applied.**

The numbers on your work have the following meanings. Note that the % figures after each comment indicate the % of students who required that particular comment.

3	Your axis is not numbered correctly. Always select chart type XY SCATTER when using MS Excel. (25%)
2	Lab. reports should have the following sub -headings and should be presented in the following order; introduction, method, results, conclusion. (88%)
5	Your graph should display the individual data points, in addition to a best-fit line. The data points should NOT be joined together by a "dot to dot" type line. (63%)
1	When comparing your result with value(s) from the literature, you should state the author, title, year, and publisher of any data sources you refer to. In this experiment $k_2 = 1.0 \times 10^{-2}$ mol^{-1} dm^3 s^{-1} (J. Chem. Ed. 1997, page 972). (75%)
4	Incorrect units/units not stated clearly. In this expe riment, t in s, V in ml, k_1 in s^{-1}, k_2 in mol^{-1} dm^3 s^{-1}, ln (V_{inf} - V) is unitless. Correct units should be stated in all column headings and on graph axes. (38%)
6	A best-fit line (BFL) is required. In a plot of ln(V_{inf} - V) versus t, the BFL should be li near. In a plot of V versus t, the BFL should be curved and should pass through the values of V_{calc}. (50%)
7	Your graph title is unclear/incorrect or absent. As a minimum, it should state the quantities plotted on the Y and X axes. (38%)
8	There is insuffi cient discussion of experimental error in your work. The main errors in this practical result from the volatilisation of I_2 during heating and uncertainties in the end point due to incomplete decolourisation of the starch indicator. (50%)

Electronic Feedback 5. Licenced to Dr. Philip Denton until 01/07/2002.

Figure 1: Example feedback sheet produced using the electronic feedback procedure.

List

	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
	e-mail		Forename	Surname	Regno	
1	PACCBAKE	MISS	CATHERINE	BAKER	66342	
2	PACJBALE	MISS	JOANNE MA	BALEED	206282	
3	PACABASS	MR	AHMED HAF	BASSI	46634	
4	PACRBERA	MISS	RAMANDEE	BERAHNEG	283599	
5	PACABULL	MR	AMIR	BULLOCK	125146	
6	PACJCAME	MISS	JULIE ELIZA	CAMERON	328365	
7	PACACAVE	MR	ALASTAIR J	CAVE	133879	
8	PACBCHAK	MR	BENJAMIN	CHAKRABO	134667	
9	PACPEVAN	MR	PARTHA PR	EVANS	628045	
10	PACHFAZL	MISS	HAYLEY	FAZLEE	234042	
11	PACMHARR	MR	MOHAMMA	HARRIS	265837	
12						

Figure 2: Example *List* sheet from the spreadsheet Feedbac.xls.

electronic details in the order favoured by their institution.

The marker enters specific details of the activity that is being assessed into the *Header* worksheet, including the title of the coursework and the maximum mark that can be awarded. The assessor can also put in statements that they wish to appear in the 'Comment' section of the feedback sheet, Figure 1. Some of these comments will only appear if the student's mark falls within certain boundaries. Thus, the example *Header* worksheet in Figure 3 has been completed so that any student awarded a mark of 60 % or above, but below 70 %,

Header

Enter filename	Uchemed
Title of Coursework	Determination of a rate constant for the reaction of I}-(aq) and S{2O{8}2}-(aq).
33	**Maximum Mark**

Maximum % Mark	**Comment**
70	Excellent work.
60	Very good work.
50	Good Work.
40	Satisfactory work.
30	Unsatisfactory work.
0	Poor work.

	Other Comments
Top Mark	Top of the class, well done!
Late Work	This work was submitted late. A lateness penalty has been applied.
All Students	The numbers on your work have the following meanings. Note that the % figures after each comment indicate the % of students who required that particular comment.

Figure 3: Example *Header* sheet from the spreadsheet Feedbac.xls.

Annotations

Determination of a rate constant for the reaction of I}-(aq) and S{2O{8}2}-(aq).
Dr. Philip Denton

Number of scripts marked (to date)	8
Highest Mark (%)	76
Average Mark (%)	46
Lowest Mark (%)	24
Standard Deviation (%)	16

	Meaning of the annotations	**% of students with this annotation**
1	When comparing your result with value(s) from the literature, you should state the author, title, year, and publisher of any data sources you refer to. In this experiment k{2 = 1.0 x 10}-}2 mol}-}1 dm}3 s}-}1 (J. Chem. Ed. 1997, page 972).	75
2	Lab. reports should have the following sub-headings and should be presented in the following order; introduction, method, results, conclusion.	88
3	Your axis is not numbered correctly. Always select chart type XY SCATTER when using MS Excel.	25
4	Incorrect units/units not stated clearly. In this experiment, t in s, V in ml, k{1 in s}-}1, k{2 in mol}-}1 dm}3 s}-}1, ln (V{i{n{f - V) is unitless. Correct units should be stated in all column headings and on graph axes.	38
5	Your graph should display the individual data points, in addition to a best fit line. The data points should NOT be joined together by a "dot to dot" type line.	63
6	A best-fit line (BFL) is required. In a plot of ln(V{i{n{f - V) versus t, the BFL should be linear. In a plot of V versus t, the BFL should be curved and should pass through the values of V{ c{a{l{c.	50
7	Your graph title is unclear/incorrect or absent. As a minimum, it should state the quantities plotted on the Y and X axes.	38
8	There is insufficient discussion of experimental error in your work. The main errors in this practical result from the volatilisation of I{2 during heating and uncertainties in the end point due to incomplete decolourisation of the starch indicator.	50

Figure 4: Example *Annos* sheet from the spreadsheet Feedbac.xls.

receives the comment, "*Very good work.*" Additional comments can be directed to those students who are subsequently identified as having handed work in late, and to the student who secured the highest mark.

Tutors are required to enter a series of feedback statements into the *Annos* worksheet, Figure 4. The statistical information that appears on this sheet will be considered subsequently. Typically, the feedback comments are those which past experience shows are most likely to be needed during marking, e.g. "*You have failed to state the correct units*". Clearly, the number of written comments required will depend on the nature of the assessment activity. When a large group of students have submitted practical reports on a particular experiment, for example, the same errors and misunderstandings crop up time and time again, limiting the number of statements that is required. Since Excel does not have the facility readily to format text, comments that contain superscripts, subscripts, line breaks or tab spaces can be entered using a series of special characters. Thus, '{' = convert next character to a subscript, '}' = convert next character to a superscript, '^' = insert line break, '¬' = insert tab space. For example, the formula of the persulphate ion is 'S{2O{8}2}-' using this system.

Marking

The three worksheets described up to this point, *List, Header* and *Annos* can be completed before the students have submitted their work. During marking itself the students' scripts are annotated with digits where each number corresponds to one of the feedback comments on the *Annos* worksheet. In this way tutors can avoid having to write the

Number	%		Cohort	Mark awarded to late work before the imposition of a lateness penalty. /33	Final mark awarded or PMC. /33	Annotations (max. 25)							
1	24	CATHERINE BAKER	A	14	8	3	2	5	1	4	6	7	8
2	PMC	JOANNE BALEED	A		PMC								
3	52	AHMED BASSI	A		17	2	3	1	7				
4	48	RAMANDEEP BERAH	A		16	5	4	1	2				
5	39	AMIR BULLOCK	A	16	13	6	8	1					
6	45	JULIE CAMERON	A		15	1	2	5	6				
7	76	ALASTAIR CAVE	A		25	2	5	6					
8	52	BENJAMIN CHAKRAB	A		17	5	8	2					
9		PARTHA EVANS	A										
10	30	HAYLEY FAZLEE	A		10	4	7	2	1	8			
11		MOHAMMED HARRIS	A										

Figure 5: Example *Numbers* sheet from the spreadsheet Feedbac.xls

same comment repeatedly on students' work. Upon receipt of their marked work, students can appreciate the meaning of the numerical annotations on their work by referring to their feedback sheet.

Before any feedback sheets can be generated it is necessary to provide the *Feedbac.xls* workbook with details of which feedback comments were assigned to which students and also the marks that were awarded. Tutors can put this information into the *Numbers* worksheet, Figure 5, which is created automatically when a class list has been entered into the *List* worksheet. Adjacent to each name are 27 empty cells. The first cell is only filled in if a student handed their work in late. Tutors enter the mark awarded before the imposition of any lateness penalty. The feedback sheet for this student will then include the comment for late work that is specified on the *Header* worksheet. Into the second cell, tutors enter the final mark that has been awarded to the work. This score is automatically converted into a percentage by dividing it by the maximum mark that has been entered into the *Header* worksheet. Alternatively, tutors can enter 'PMC' into the second cell if it is known that the student is not going to submit any work because they have personal mitigating circumstances. The calculated % mark or 'PMC' is displayed on the *Numbers* worksheet in the second column. In the remaining 25 blank cells that are adjacent to each student's name, tutors put the numbers that correspond to the feedback statements that have been allocated to that class member on their marked script. Thus, the maximum number of comments that can be assigned to a particular student is 25.

Upon completion of the *Numbers* worksheet, the *Annos* worksheet displays statistical details relating to the activity, Figure 4. Thus the maximum, average and minimum percentage marks are reported and this information is reproduced on the feedback sheet, Figure 1. The *Annos* worksheet also computes the percentage of students who required a particular comment during marking, Figure 4. These values are reproduced on the feedback sheet as a percentage figure in brackets after each comment, Figure 1.

Generating and returning feedback to students

When a mark for a particular student is entered into the *Numbers* worksheet, the spreadsheet automatically generates the corresponding feedback sheet. Excel does not have the capability to produce large amounts of formatted text. Thus, before printing or e-mailing, the feedback sheets must be copied and pasted into the MS Word 97 document *Fb.doc*. Both *electronic feedback* programmes, *Feedbac.xls* and *Fb.doc*, incorporate a series of visual basic programmes that enable this copying, pasting and formatting procedure to be accomplished via a couple of mouse clicks.

Evaluation of electronic feedback

The educational benefits of electronic feedback were evaluated by studying the frequency with which feedback comments that relate to a particular error were used during marking. Clearly, one would hope that the frequency of use of such comments would gradually decrease over time as students reacted to their feedback and corrected their mistakes.

The attitudes of students to the electronic feedback strategy was ascertained by their responses to a structured questionnaire that was completed by 58 first year undergraduate students within the JMU

Meaning of the annotation (abridged)	% of students with this annotation (7/10/99)	% of students with this annotation (23/9/99)
Lab. reports should be presented in the following order; introduction, method, results, conclusion.	60	36
Your graph axis is not numbered correctly. Always select chart type XY SCATTER when using MS Excel.	36	2
Incorrect units/units not stated clearly.	98	74
Your graph should display the individual data points, in addition to a line.	34	2
Your best-fit line is incorrect and/or absent.	91	31
Your graph title is unclear/incorrect or absent.	57	7
Your graph axis is not labelled correctly or is not labelled at all.	21	0

Figure 6: Assessment profiles from two first year undergraduate chemistry practicals.

School of Pharmacy and Chemistry. This was in addition to three focus groups, each consisting of three students chosen at random. Members of staff were also requested to offer their views on the software after it was presented to them during a JMU training session.

Results

The *electronic feedback* method has been used to assess physical chemistry laboratory reports and worksheet assignments submitted by first and second year undergraduates at JMU. The procedure has been found to work well in practice. Ideally, the bank of feedback statements should be written before the assessor receives the students' work to enable the marking to be completed as quickly as possible. There is no reason, however, why it cannot be gradually built up during marking itself.

When marking laboratory reports, the same initial bank of general feedback comments can be used. These are then edited and augmented so that they are appropriate to each experiment. Typically, about 25 distinct comments are required. Many of the comments relate to core skills such as report writing and the graphical representation of experimental data. The frequency with which particular comments were used when marking two first year undergraduate physical chemistry experiments, conducted two weeks apart, is shown in Figure 6. By the time the students came to undertake the second practical they had already received e-mailed feedback on the first. As is evident, the ability of the students to present their work in an appropriate scientific manner had improved markedly over this period.

Annotating students' work with numbers in place of comments ensures that marking is relatively straightforward and rapid. Moreover, the *List* and *Header* worksheets of the *Feedbac.xls* file can be filled efficiently if electronic class list information is available. The time taken to complete the *Annos* worksheet will depend on the number and complexity of the feedback comments. As discussed above, however, if the assessed assignment is similar to one that has been set previously, tutors may find that it is possible to use an existing bank of feedback comments that has been appropriately modified. In this way, the time taken to enter the *Annos* worksheet can be considerably reduced. The *Numbers* worksheet can be completed quickly if marked scripts are first arranged in class list order, before the marks and the numerical annotations that appear on the work are entered.

Students reacted positively to the electronic feedback procedure when questioned in the focus groups and in responding to the questionnaire. All the interviewed students felt that the e-mailing of feedback was an efficient way to receive details of their performance in an assessment, as it removed the requirement for them to wait until the next time they met the lecturer. It was evident that students were comfortable with the principle of receiving feedback when they were at a computer terminal on their own. The focus groups confirmed that they were more likely to pay attention to the electronic feedback that is returned quickly and felt that it was then acceptable to wait 2-3 weeks for marked coursework to be returned.

The questionnaire revealed that 88% of the undergraduates felt that it was useful to have written feedback e-mailed to them in advance of receiving their marked script. The vast majority stated that they appreciated knowing the maximum, average and minimum marks for the activity (91 %) as well as their position in the class (88 %). All questioned students found the comments on their feedback sheet useful and most of the class (81%) felt that they had received more written feedback than they normally obtained from tutors. It became clear that one of the particular advantages of the electronic feedback procedure is its ability to return lengthy, detailed comments on a particular aspect of the assessment. The responses of two students were typical, "*It is a helpful method of marking as it enables you to see how and why mistakes were made...*" and, "*It offers a more in-depth description of how you have gone wrong*." A number of students also commented that the printed feedback sheets overcame difficulties associated with the legibility of staff handwriting. In response to the question, "*should electronic feedback be used more regularly within the School*", 100% of respondents said, "*Yes*".

After the staff training session, 31 colleagues returned written sheets to provide feedback on the session. Those staff who have a familiarity with Excel reported minimal difficulties using the software. One member of staff commented, "*A fairly complex piece of software which I will feel more confident of using once I've tried it out using my own annotations. Educationally, a very sound method*." Other staff acknowledged that the procedure could become second nature with practice. 5 members of staff said they would definitely not use the software in future, either because they had an existing electronic system that they preferred, or because they had experienced difficulties using the software.

Discussion

Up to now, the electronic feedback technique has been used primarily in the grading of chemistry coursework. The procedure is quite general, however, and can be used for any assessment where it is expected that students will make the same errors repeatedly. It is clear that the electronic feedback approach can make the marking process considerably less onerous as it removes the requirement to annotate students' work with repeated hand-written comments. The package would be of particular interest to those tutors who find that, using conventional methods, they are unable to return as much feedback as they would wish to. Although this approach is perhaps less personal than traditional marking, there is no reason why tutors cannot supplement their printed feedback with hand-written comments to individual students.

The two files that comprise electronic feedback are password protected to prevent the accidental overwriting of essential subroutines. Thus, there are limited opportunities for customisation. Tutors can have some control over the final appearance of the feedback sheets, however, and may choose to omit details of the maximum, average and minimum marks as well as the student's position in the class. Tutors who prefer to allocate grades instead of marks can choose to hide the percentage marks that are normally displayed on the feedback sheets and can write feedback statements such as 'Grade B+'. Each statement can have a particular number associated with it and these can be allocated to students in the customary manner.

The software need not necessarily be used exclusively when marking assessments where the same errors crop up repeatedly. If they so wish, tutors can write a single, lengthy feedback comment for each student and use the software to generate the corresponding feedback sheets. In this way the software can be used to e-mail feedback to students on highly individual assessments such as project work. If this approach is adopted, the number that corresponds to the feedback comment, and the percentage of students requiring that statement, may be omitted from the feedback sheet.

Acknowledgement

This work was completed as part of the Postgraduate Certificate in Learning and Teaching in Higher Education 1999/2000. I wish to thank all the tutors associated with this course.

Supporting material

Copies of the requisite software and a user guide are available from the author. Interested persons should send a stamped addressed A4 envelope and a formatted 3½" disk. Please include your e-mail address so that you may be contacted subsequently for your opinions of the software. Respondents should indicate how they wish their title and name to appear on the feedback sheet, as this information cannot be changed subsequently. This is a security precaution that is included so as to prevent the unauthorised proliferation of the software.

References

1 G. Brown, J. Bull and M. Pendlebury, *Assessing Student Learning in Higher Education*, Routledge, London, 1997, p. 7.

2 G. Gibbs, *Improving the Quality of Student Learning*, Technical and Educational Services, Bristol, 1992, p. 17.

3 G. Gibbs and T. Habeshaw, *Preparing to Teach: An Introduction to Effective Teaching in Higher Education*, Technical and Educational Services, Bristol, 1993, p. 95.

4 P.Ramsden, *Learning to Teach in Higher Education*, Routledge, London, 1993.,

5 T. King, *Using an Optical Mark Reader for Continuous Student Assessment: a Case Study in Higher Education in Computer Assisted Assessment*., ed.-s, J. Darby and J. Martin, 1994, p. 23.

6 G. Gibbs, S. Habeshaw and T. Habeshaw, *53 Interesting Ways to Assess Your Students*, Technical and Educational Services, Bristol, 1986, p. 77.

7 B.S. Nicholls, *U. Chem. Ed.*, 1998, **2**, 10.

8 B.S. Nicholls, *U. Chem. Ed*., 1999. **3,** 22.

9 S.L. Helgeson and D.D. Kumar, *Journal of Computers in Mathematics and Science Assessment*, 1993, **12**, 227.

10 L. Barnett, D. Brunner, P. Maier and Warren, *A Technology in Teaching and Learning: A Guide for Academics*, Interactive Learning Centre, University of Southampton, 1996, p. 59.

11 S. Habeshaw, G. Gibbs and T. Habeshaw, *53 Problems with Large Classes: Making the Best of a Bad Job*, Technical and Educational Services, Bristol, 1992, p. 136.

Experiencewith a Random Questionnaire Generator in the chemistry laboratory and for other continuous assessment

Paper

Mary R. Masson
Department of Chemistry, University of Aberdeen, Meston Walk, Aberdeen AB24 3UE.
E-mail: m.masson@abdn.ac.uk

The Random Questionnaire Generator, a suite of programs designed to produce randomised multiple-choice tests for assessment of a first year chemistry class, has now been in use at Aberdeen University for three years. It has proved popular with students and staff and gives a much more reliable mark for each student than the previous system. The Creator program has also been used to generate tests for use in continuous assessment tests for students at Level 2.

Introduction

At Aberdeen University, as at many of the other older Scottish Universities, a large proportion of B.Sc. students take the Level 1 Chemistry course. There is no separation of intending chemists from other science students. All students taking the first year course follow the same laboratory course, which requires attendance at one 3-hour class each week. Several set experiments run on each lab day, and students rotate around these according to a pre-defined rota.

Students make records of their work in their laboratory manual during the laboratory class. The records include calculations, graphs, data analysis and answers to questions. Until three years ago all laboratory manuals were checked and marked by a member of staff before the student was permitted to leave the laboratory. The increase in student numbers during the 1990s meant that there could be up to 120 students in the lab during any lab session. This overloaded the system, caused unacceptable queuing and resulted in some students stopping lab work early so as to get to the front of the queue (a practice which was quickly copied by others).

Staff identified the following problems with the marking process:

- It was impossible to award a meaningful and consistent mark in the time available to assess each student (less than 2 minutes) and as a result the marks did not discriminate between able and less able, or even between conscientious and careless students.
- Opportunities to teach through interaction with students at the bench were significantly reduced by the time spent on marking.
- Student time spent on laboratory work was unacceptably reduced not only by the need to queue, but also by the tendency to stop work early.

In considering how to reorganise marking procedures to alleviate these problems we concluded that a computer-based test offered the most promising way forward. We identified the following characteristics of a satisfactory assessment procedure:

- Each student would be provided with an individualised test to be completed by the end of the laboratory session; this individualisation would both prevent simple copying of answers and allow the test to take account of the fact that in many experiments students are given different samples to analyse.
- The questions would be worded in a friendly way so that students would recognise the test as a valuable learning experience.
- Questions would attempt to ensure that the information provided in the laboratory manual is read and understood (preferably in advance of the class) and would consolidate the theoretical background by asking about new terms and definitions introduced in the experiment, and understanding of general information related to the experiment.
- The test would be designed to provide a mark for the recording of observations made during the experimental work (e.g. the colour of a precipitate formed at a particular point), for the quality of the results obtained and would give practice in calculations with dummy data.
- When dealing with observations and results, it should be possible to award fractional marks.
- As far as possible the test would be marked automatically using an optical mark reader.

Our Department has been using multiple-choice testing in examinations since the early 1960s and has followed with interest the literature on the requirements for the design of effective questions. This has been revisited in recent years as a result of the increase in popularity of computer-based assessment.[1, 2] Also, we have noted since we started our project that other institutions have reported the introduction of computer-based pre-lab and post-lab tests.[3, 4] However none of the published work appears to have been concerned with the provision of a single test based on a particular laboratory class in which individual questions are designed to be completed at different times throughout the class. We looked at other testing systems that include randomised selection of questions (for example Turton[5]), and at commercial systems such as Question Mark Designer,[6] but found that these were designed for computer delivery of tests, and did not offer any possibility of customisation to meet all our needs. We therefore decided that the best option was to design our own style of test.

The system we designed has now been in operation for three years.[7,8]

Methods

Creating the questions

The first year laboratory course consists of 20 sessions in each of which each student completes a different experiment. Before leaving the laboratory each student completes a test of 20 questions and hands in a coded answer sheet. Each test paper is unique to the student concerned and is generated on paper for each student from a bank of multiple-choice questions. Students collect their test papers at the start of each laboratory session, and answer the questions at appropriate points during and at the end of the experimental procedures. Because all the tests are different, we are able to allow students to discuss the answers to questions. Students are also encouraged to seek help from demonstrators with questions they find difficult.

The answers to the test questions are entered directly onto a copy of the University's standard, machine-readable form for marking by an optical mark reader; this form simply lists the question number and offers the choice of five boxes corresponding to the alternative responses provided (five boxes are always provided even though some questions are only provided with three or four responses). Individual test papers, each of which is prepared as a Microsoft Word document, are created by the computer from the bank of questions. Typically the question bank for each experiment consists of 24 different basic questions, but the variety is considerably increased because many of these basic questions have a number of variants (2 – 8). Variants are generated in several different ways. Sometimes it was possible to devise equivalent questions that were totally different, or (in the case of dummy calculations) to provide different data for the same calculation. Sometimes it was possible to use a fixed set of responses, but different question stems, one corresponding to each response. However, in some cases, the only acceptable variation was to change the order of the possible responses. With such questions, it is obviously easier for students to collaborate, but we were satisfied that they would at least have to read the question and its responses carefully — it would not be possible just to find out from a neighbour that the answer to question 3 was B.

The provision of a machine-readable test of observations made and of results was tackled in two ways. The first is used in experiments where all students should in theory obtain the same answer. The style of question used here is:

- **From the calibration graph, in which of the following ranges was the concentration of potassium ions in the riverwater?**
  A Less than 8 mg l^{-1}
  B 8–9 mg l^{-1}
  C 9–10 mg l^{-1}
  D 10–11 mg l^{-1}
  E More than 11 mg l^{-1}

We tried here to avoid making the middle response the correct one, so that students could not use our ranges to guess the answer that we were expecting. The evidence we have from student queries is that they are very keen to code their results correctly, rather than attempt to cheat the system. The correct response is awarded a full 1 mark, but other responses may be awarded partial marks, based on our knowledge of the errors of the experiment.

The second method is required for experiments where students are expected to get different results, and where the mark is to be awarded for aspects of experimental work that involve human judgement. For these questions, the student is informed what the marks are awarded for but is instructed not to provide an answer. Instead, responses to be coded on the machine-readable answer sheets are decided by demonstrators after examining the laboratory manual and checking calculations where necessary. This is done throughout the class and does not usually cause delays. Examples of the these questions are:

Has the mass of the iron compound been recorded correctly?
Are the titration volumes in good agreement?
Has the percentage yield been calculated correctly?
Has the graph been drawn neatly and correctly?

Demonstrators have a key for each experiment, which details the appropriate response for these questions, which we define as 'demonstrator-marked'. For each question, the key provides a letter that corresponds to the full mark of 1; the other letters correspond to fractions of a mark. Thus, if the key letter is C, students awarded response C receive a mark of 1, but other responses receive 0.75, 0.5, 0.25, or 0. The letters corresponding to the correct result are different for the different experiments, are kept secret from the students, and are changed from time to time. Amongst other characteristics of student work, these demonstrator-marked questions require demonstrators to check calculations, to examine graphs for neatness and correct plotting, and to examine the appearance of organic products. Marking guidelines leave as little room as possible for personal interpretation so that postgraduate demonstrators are able to assist the staff with marking. Demonstrator-marked questions are included in the same format and in the same place on all the test papers for a particular experiment. They are therefore designated as ‘fixed’.

Questions dealing with other aspects of the experiment (for example, calculations with dummy data, and questions dealing with background theory) are of a more conventional style. For the purposes of creating test papers these questions are designated as ‘fixed’ (one of the variants appears in every test paper, always in the set position), ‘compulsory’ (one of the variants appears in every test paper, but in any available position) or as ‘optional’ (a question which need not be selected).

Depending on the experiment, a test paper will contain 3–20 fixed questions, 0–10 compulsory questions, and 17–20 optional questions selected from a bank of up to 30 (basic questions) with a total of up to 80 variants.

Creating the test papers

The test papers are created from the question banks and the corresponding Questionnaire Definition Files, by using the Creator program, which is written in *Delphi* (like the others in the suite). The program reads a set of daily *Excel* spreadsheets (giving lists of student names, laboratory numbers, and class codes), another spreadsheet giving the rota of experiments (which relates student laboratory numbers with experiments for each week), and a file, which defines all the experiments. The program uses the student’s laboratory number together with a ‘day code’ defining the day of the week and the week number to set the rules by which the algorithm selects the questions for each student in a manner that appears to be random. Each test will include all the fixed questions, in their defined positions, along with the Compulsory questions and enough Optional questions to give a total of 20; the positions of the “c” and “o” questions are different for each different test. The algorithm also determines which of the optional questions are selected and (for questions with variants) it also determines which variant is selected.

To simplify the handling and distribution, the font size and margins for the questionnaire are set to permit each test to be printed on a single sheet of paper. An example of a test is shown in Figure 1. Once the student names are available, laboratory numbers have been allocated, and the daily *Excel* spreadsheets prepared, questionnaires are simple to produce. Normally they are generated for all students for a given day and week number in a single batch (although if necessary a single form can be printed). This is repeated for other days and week numbers. The limiting factor in the process is the printer — but with a fast printer, tests for 20 lab sessions (for 5 weeks, with 4 lab classes each week; up to 3000 tests) can be printed in a single day.

Marking program

The laboratory technicians scan the completed machine-readable forms by using a Scanmark 2000 (http://www.scantron.com/scan/sm2000.htm) optical mark reader (omr). The machine is set to reject forms if marks are missing or not dark enough to be readable; it can also reject some invalid codes. Any rejected form is returned at once to the student, with advice to make marks darker, or insert or amend codes and then the form is rescanned. We bought our own reader for this project, but use the University's standard printed forms. The omr software writes the data to an ASCII text file.

Each week, the four daily files are processed by the Marker program, which uses the same algorithm as the Creator program to determine, for each student, which questions have to be marked. It then checks the answers in the answer file, and awards the appropriate mark. It has not proved possible to make the marking process fully automatic, because students sometimes make errors in entering their laboratory numbers and/or day codes, but the program attempts to flag these, and the flagged entries are corrected manually. When all corrections have been made, the program is instructed to insert the marks into the Excel spreadsheet, which also serves as the register for the class. At the time of marking, the responses are recorded in a text file for later analysis, if desired, for example to calculate the fraction of the class with the correct answer to every variant of every question (the facility value).

A special arrangement had to be made for the "Unknown Samples" that students have to identify in the Inorganic part of the laboratory class. There are 47 different samples, so it was not considered necessary to generate fully randomised tests.

2071 JOHN SMITH Cations part 2 Day code: 0321

Q1. The colour of the solid and the stock solution of compound T is?
A Colourless / white B Pink / red C Green D Blue

Q2. With which of the following pairs of cations are the colours of the solid and stock solution of compound T compatible?
A Cu^{2+} and Cr^{3+} B Cu^{2+} and Fe^{2+} C Cr^{3+} and Co^{2+} D Ni^{2+} and Fe^{2+}

Q3. On heating the precipitate a change in colour was noted. What was the final colour observed?
A Blue B White/colourless C Green D Black

Q4. Which of the following is the only cation to fit the result obtained when sodium hydroxide solution (dilute then excess) was added to the compound?
A Ni^{2+} B Cu^{2+} C Cr^{3+} D Fe^{2+}

Q5. Unknown T was copper sulphate 5-hydrate. Which of the following is the formula for this compound.
A $(CuSO_4)_5\,H_2O$ B $Cu(SO_4)_2.5H_2O$ C $CuSO_4.5H_2O$ D $Cu_2SO_4\,H_2O$

Q6. Which of the equations below represents the formation of the precipitate of copper hydroxide?
A $Cu^{2+}(aq) + 2OH^-(aq) \rightarrow Cu(OH)_2(s)$
B $2Cu^{2+}(aq) + OH^-(aq) \rightarrow Cu_2OH\ (s)$
C $Cu^{2+}(aq) + OH^-(aq) \rightarrow CuOH(s)$
D $Cu^{2+}(aq) + 2OH^-(aq) \rightarrow CuOH_2\ (s)$

Q7. What is the formula of the compound formed on heating copper hydroxide?
A CuS B CuO C CuOH D $Cu(NO_3)_2$

Q8. The precipitate redissolved on adding excess ammonia solution owing to the formation of a complex. Is the complex
A An anion B A cation C A neutral molecule

Q9. What is the name given to this type of complex?
A Hydroxo B Ammine C Amine D Hydrate

Q10. Which of the following equations represents the formation of the copper complex ?
A $Cu^{2+}(aq) + 6NH_3(aq) \rightarrow [Cu(NH_3)_6]^{2+}(aq)$
B $Cu^{2+}(aq) + 4NH_3(aq) \rightarrow [Cu(NH_3)_4]^{2+}(aq)$
C $Cu^{2+}(aq) + 4NH_3(aq) \rightarrow [Cu(NH_3)_4]^{4+}(aq)$
D $Cu^{2+}(aq) + 6NH_3(aq) \rightarrow [Cu(NH_3)_6]^{6+}(aq)$

Q11. On addition of excess of concentrated hydrochloric acid to a solution of unknown T a colour change was observed. Which of the following changes in colour best fits your observation?
A Blue to greenish yellow B Blue to violet C Blue to red D Green to blue

Q12. The change in colour is due to the formation of a complex. Is the complex
A An anion B A cation C A neutral molecule

Q13. What is the name given to the type of complex formed with hydrochloric acid?
A Hydroxo B Chloro C Amine D Hydrate

Q14. Which of the following represents the formula of the complex formed?
A $[CuCl]^-$ B $[Cu_2Cl]^{2+}$ C $[CuCl_4]^{2-}$ D $[Cu_2Cl_4]^{2-}$

Q15. Which of the following is the correct formula for the precipitate formed when sodium sulphide was added to T?
A CuS_2 B Cu_2S C CuS D Cu_2S_2

Q16. From the colour of compound U and its solution, which of the following groups of cations can you say are definitely not present in the compound?
A Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Cd^{2+}
B Cr^{3+}, Fe^{3+}, Co^{2+}, Ni^{2+}, Cu^{2+}
C Zn^{2+}, Al^{3+}, Pb^{2+}, Sn^{2+}, Sn^{4+}
D Cd^{2+}, Mn^{2+}, Fe^{2+}, Ag^+

Q17. Which of the following groups of cations is compatible with the result from the reaction between compound U and sodium hydroxide solution?
A Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Cd^{2+}
B Cr^{3+}, Fe^{3+}, Co^{2+}, Ni^{2+}, Cu^{2+}
C Zn^{2+}, Al^{3+}, Pb^{2+}, Sn^{2+}, Sn^{4+}
D Cd^{2+}, Mn^{2+}, Fe^{2+}, Ag^+

Q18. Which one of the following cations can be eliminated because it would have given a precipitate if the solid was tested with nitric and sulphuric acids?
A Ca^{2+} B Mg^{2+} C Cd^{2+} D None

Q19. Unknown U was magnesium sulphate 7–hydrate.Which of the following is the formula for this compound?
A $(MgSO_4)_7\,H_2O$ B $Mg(SO_4)_2.7H_2O$ C $MgSO_4.7H_2O$ D $Mg_2SO_4\,.7H_2O$

Q20. Which of the following is the formula for the precipitate formed when sodium hydroxide was added to a solution of U?
A $Mg(OH)_2$ B $Mg_2(OH)_3$ C MgOH D Mg_2OH

Figure 1. An example of a questionnaire

Instead, a limited number of tests were prepared for each of the samples used. The pre-prepared questionnaires are associated with the samples rather than with individual students. These tests are identified to the Marker program by special codes printed on the test forms in place of the normal day codes. Different students complete different numbers of unknown samples, so all the marks achieved by one student are summed in a single spreadsheet cell.

Providing students with feedback

Feedback to the students is provided in two ways. In the first, marks are made available by the Laboratory Mark Reader program, which runs on two low-specification computers in the laboratory. This program reads copies of the main Excel spreadsheets, and allows students to see the mark achieved in the previous week(s), and also to find out the numbers of the questions they have answered incorrectly. Students are advised to retain their questionnaires so that they can review these questions; they are advised to consult a demonstrator if they do not understand their error.

The second form of feedback is provided only to students who get marks below a given threshold (usually 12 out of 20). These students are offered a printed report showing their incorrect responses and they are particularly recommended to consult a demonstrator for advice about their errors. The printouts include the full text of questions answered incorrectly, but questions related to the student's experimental data are normally labelled in the answer file so that they are excluded from these reports. (The same applies to the first form of feedback.)

Results

The facility value calculated for each question (and each variant) provides evidence that no questions appear to be either so easy or so difficult that they provide no useful discrimination between students. Typically, the facility values fall between 0.6 and 0.8, which we regard as satisfactory here. In a normal exam, an 'ideal' question would have a value of 0.5, because half the class got the question correct; but a test should include a range of facility values in order to discriminate across a range of student abilities (1).

Furthermore, the facility values for variants of the same basic question are essentially the same. These results provide reassuring evidence that the tests are discriminating effectively. A point of particular interest arises from our examination of the facility values for questions, which differed only in the order of the distractors. We observed considerable variation in the numbers choosing the various wrong answers, but we found only two for which the correct answer had a higher facility when the

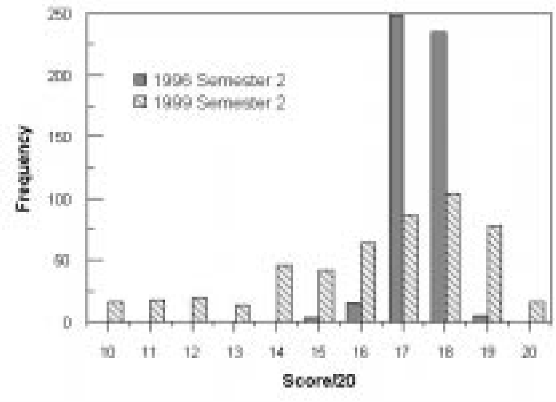

Figure 2. Comparison of laboratory marks awarded by the old and new systems

nearest distractor immediately preceded it. In view of the large number of questions we examined, this was not statistically significant and our observations are therefore inconsistent with the suggestion that the relative locations of the correct answer and 'nearest' distractor can have a significant effect on the facility of questions.[9] A useful reminder that 'objective testing' is not fully reliable and reproducible was provided by a question with only 3 responses (A, B and C), but for which 6% (1998) and 4% (1999) of students marked response D.

Figure 2 shows a comparison of the marks awarded during a single semester (1999, semester 2) using the new system with those awarded during the last semester in which the old system was used (1996, semester 2). The distribution of marks is still skewed towards the top end of the marking scale, but the new system has resulted in a much fuller use of the total mark scale. Table 1 provides a comparison of the marks obtained in all the semesters since 1996 (2) and shows that 1999(2) is typical of semesters since the introduction of the new system. (Although there is no statistical justification for calculating a standard deviation for data, which are clearly not drawn from a Normal Population, it is a convenient way of providing a crude comparison between semesters).

Semester	96/2 (Old)	97/1	97/2	98/1	98/2	99/1	99/2
Mean	17.7	16.0	15.7	16.0	15.9	16.5	16.2
Standard Deviation	0.5	2.6	1.8	2.5	3.0	3.0	2.7

Table 1. Statistics for Laboratory Marks

Using the new system, the distribution of marks for laboratory work is similar to the distribution for closed examinations. This is illustrated in Figure 3, which shows a plot of lab marks *vs.* exam marks for

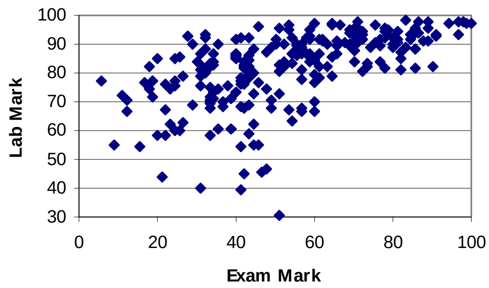

Figure 3. An example of correlation between laboratory and examination marks

semester 1999(2). Almost every student obtained a higher mark for lab work than for the closed examination. However the correlation coefficient is highly significant ($R = 0.57$, $n = 243$, $P < 0.001$).

We are satisfied that the distribution of marks obtained with the new system is a better reflection of our students' performance in the laboratory classes. We find it hard to accept that the data shown in Figure 2 for semester 1996(2) really mean that all our students truly deserve marks in the top quartile of the scale. And we are neither surprised nor concerned that most students achieve a higher mark for lab work than they do in examinations because it is possible in the laboratory to compensate for mistakes or lack of aptitude by perseverance and hard work. Furthermore, we have noted a marked improvement in student attitude to laboratory work since the introduction of the new system; the great majority actively strive to achieve high marks and keen to learn how they have lost marks, to the extent that there have been many requests for the feedback printouts to be made available for all students. In official Student Course Evaluation Forms, over 90% of respondents said that they thought the practical work was well organised. In addition, around 85% said they found the laboratory work interesting (in contrast to Johnstone's report[10] that the majority of young lecturers interviewed by him had, as students, found labs boring).

Staff and demonstrators also report favourably on the new system. The laboratory is now able to close at the advertised closing time. The automation of the recording of attendance and mark has significantly reduced the administrative work associated with the running of the class.

Discussion

There can be no doubt that we achieved our first objective of removing the problem of queues with the consequent advantage that the students do not stop work earlier than is justified by their progress. There are other advantages. The system for marking and recording of marks operates very efficiently giving demonstrators more time to teach in the laboratory. The use of paper tests means that, for some kinds of experiment, we can completely integrate the tests with the experiments so that they are in-lab rather than post-lab tests. The Master *Excel* spreadsheet allows us to monitor student attendance and take action if necessary. Our comparisons of the new lab marks with old ones and with examination marks have convinced us that we now have a much more reliable mark for the laboratory work than we did in the past so that we now have no worries about the weighting of 20% given for this mark in the end-of-module degree examinations, and indeed we are giving thought to increasing this.

The original version of the Reader program gave only the mark, and comments from students obtained from course evaluation forms resulted in many requests for the feedback printouts to be made available for all students instead of just those who obtained low marks. We take this to be a positive indication of a genuine wish to improve, and to comply with this request, we have very recently introduced the enhanced version of the Reader program, which informs students of the questions that they answered incorrectly.

The suite of programs we have developed is generic in that it can be readily adapted for a variety of uses. This is illustrated by our use of the Questionnaire Generator Program to create homework assignments for level 2 students doing the analytical lab course. Because this is timetabled for the end of the session, students face a conflict between their perceived need to revise and the requirement to prepare lab reports. It was therefore decided to separate the calculations from the experimental work, and issue the calculations early in the semester. We had not felt able to do this previously because of the prevalence of copying when all students receive the same homework tests, but the opportunity to create randomised tests overcame this objection.

For each experiment, six sets of "good" but not perfect data were selected from those obtained by students in previous years, and the Creator program was used to generate randomised tests for each student. There was no intention of using machine marking, so no multiple-choice responses had to be created. The students are no longer required to prepare lab reports at the end of term, because the evidence of their ability to carry out the calculations is measured from their calculations using dummy data, and their actual data is entered into a Visual Basic program which calculates and records the final answers for subsequent assignment of marks for accuracy. The course evaluation forms show that this is appreciated.

As far as the staff are concerned, there is a considerable saving in workload. Although the marking of the calculations is done manually with the help of a key indicating the data given to each student, this is much faster than marking real laboratory data because the answers are already known. More time is saved by the automated calculation of results even though it requires some human intervention, since there is no need to check the calculation itself. This procedure also serves to prevent students from trying to "fudge" their results.

Our system is an example of the power of computers in teaching with no pretence at being Computer Assisted Learning. We see this as a very positive aspect since we have long observed that the majority of our students are not very enthusiastic about the use of computer-assisted learning.

Acknowledgements

We are grateful to the University Information Services Committee for providing the original funding which allowed us to develop the database of questions and the software, to Dr Simon Heath of CLUES for provision of programming support, and to Aberdeen University Learning Technology Unit for support for recent enhancements to the software.

The programs that make up the Random Questionnaire Generator suite are completely generic in nature, although they do require the user to be running Windows 95/98/NT, together with Microsoft *Word* and *Excel*. We are willing to make them available at a modest price to anyone who is interested in using them.

References

1 S. Heard, J. Nicol and S. Heath, *Setting effective objective tests* MERTaL Publications, Aberdeen, 1997.

2 P. Race Changing assessment to improve chemistry learning, Project Improve, 1997.

3 B.S. Nicholls, *U.Chem.Ed.*, 1999, **3,** 22.

4 P.W.W. Hunter, *U.Chem.Ed*., 2000, **4,** 39.

5 B.C.H. Turton, *ALT-J*, 1996, **4**, 48.

6 *Question Mark Designer*, Question Mark Computing Ltd, http://www.qmark.com/

7 M. Masson, and W. Lawrenson, *Computer-aided laboratory assessment: in Research in Assessment XIV: Aspects of First Year Higher Education*, RSC, London, 1999, pp. 9–11.

8 M. Masson, W. Lawrenson, G. Cowan and S. Armstrong, *Random Questionnaire Generator for objective tests for laboratory practicals v. 1.1* MERTaL™ Courseware, University of Aberdeen, Aberdeen, 1997, ISBN 187315447 X

9 J. Handy and A.H. Johnstone, *Educ.Chem.***,** 1973, **10,** 47.

10 A.H. Johnstone, *U.Chem.Ed.*, 1998, **2,** 35.

Student misconceptions of the language of error

Paper

Jane Tomlinson, Paul J Dyson and John Garratt

Department of Chemistry, University of York, Heslington, York, YO10 5DD
E-mail: jlt7@york.ac.uk

We have collected student responses to questions designed to establish their understanding of twelve terms used regularly when working with error and uncertainty in quantitative data. The terms are 'reproducible', 'precise', 'accurate', 'sensitive', 'random error', 'systematic error', 'negligible', 'significant difference', 'qualitative procedure', 'quantitative procedure', 'correlated data', and 'transforming data'. In most cases less than 50% of the sample of first year chemistry students provided evidence of 'some or good understanding'. We suggest that their misconceptions are most likely to be rectified by persistently challenging the students to make explicit use of key words and concepts such as these whenever they present reports of quantitative data collected during practical work.

Introduction

One of the skills that chemistry graduates are expected to acquire is the *"ability to interpret data derived from laboratory observations and measurements in terms of their significance and the theory underlying them"*.[1] Effective data interpretation involves dealing with errors and uncertainty in the measurement of physical quantities. This is an area that requires a particular use of language. Even the word 'error' is a source of confusion to many students since they commonly regard 'errors' as personal mistakes[2] rather than recognising that *"every physical measurement is subject to a host of uncertainties that lead to a scatter of results"*.[3] Another simple example is that many dictionaries give 'accuracy' as one meaning of 'precision'[4] whereas these two words have distinctly different meanings in the context of scientific measurement.

We recently published evidence that first year chemistry students have not developed an effective use of the language used to handle error and uncertainty.[5] The majority of a sample of 65 students believed that the **accuracy** of their results would be improved by using an objective rather than a subjective method to fit a line to their experimental data, and that rather more than half of them showed confusion over the meaning of the term 95% confidence limits. In retrospect, it seemed to us that these misconceptions were easy to reconcile with their previous experiences; they are accustomed to teachers looking for a high level of 'accuracy' and consequently they readily assume that a 'line of best fit' means that the line gives the most accurate result; their use of 'confidence' in every day language is sufficiently different from its technical usage in error analysis that this can easily be a source of confusion.

The constructivist view of learning[6] leads to the expectation that it is not easy to bring about a reconstruction of a misunderstood concept already embedded in the mind.[7] This may help to explain the conclusion that *"students find it difficult to grasp the value and purpose of statistical procedures"*.[8] We therefore decided to explore student understanding of key words and concepts used in dealing with error and uncertainty in measurement. We hoped this would lead to better understanding of the concepts held by first year students, and that this would help us to devise opportunities for learning that would lead to better understanding of those issues generally considered to be important. We had the opportunity to include a set of questions as part of a first year laboratory course on Analysis taken by the same cohort of chemistry students who, in the previous term, alerted us to the possible problem.[5] At this stage they had received virtually no relevant instruction in this area, at least not since they left school. Our intention was therefore not to evaluate the course itself, but to gain a better understanding of the concepts related to errors and uncertainty, which these students brought to the course. We expected that this would help us to devise opportunities for learning that would address specifically any widely held misconceptions. We report here on the responses received to this set of questions.

Methods

In devising our questions we first attempted to obtain (through discussion with a number of concerned colleagues) a consensus view on the vocabulary with which first year students are expected to be familiar. Some of these (such as 'accuracy', 'precision', 'systematic and random error') are routinely defined or described in many standard treatments of error (e.g. ref.-s 3 and 12), others (such as 'sensitivity', 'negligible') are rarely dealt with in such texts though they are routinely used by scientists and have a special meaning in context.

We attempted to word the questions in a way that would distinguish between ability to provide a definition (declarative knowledge) and an ability to use words and concepts (procedural knowledge). Several drafts were required before we were satisfied with the set of questions. An important requirement was that we could ourselves prepare, for each question, a short answer that we regarded as demonstrating an adequate understanding of word usage in the context of analysis and error. This requirement caused us to make significant revisions to the wording of our first draft.

Figure 1 shows the final version of the set of five questions; these cover 12 key words or concepts (counting 'significant difference' and 'insignificant difference' as a single concept). Also shown is the explanatory preamble which draws attention both to the possible differences between the technical and every day meaning of some words, and to the fact that we were looking for each student's view of how to **use** the words, given that the meaning may vary with the context. The whole fits conveniently on to one side of a sheet of A4 paper.

The students in our survey were in the second term of their chemistry degree course. Each of the 103 students of the first year cohort received their own copy of the question sheet at the beginning of the six-week course on Analysis that started in the first week of the spring term of 2000. The question sheets were handed out by the Course Organiser, who gave a short verbal explanation to reinforce the points made in the preamble. No responses were received before the end

Questions set to first year chemistry students

The Language of Analysis and Error

Analysis usually involves measuring quantitatively or qualitatively one or more constituent of something. In order to communicate analytical results (including information about the effect of experimental error on their reliability), chemists use words which have technical, or specialist meanings. Many of these words are also in ordinary use; for example *accuracy, precision, random, systematic, significant.* The purpose of this exercise is to give you an opportunity to think about and explain or describe how **you** would use, in a scientific context, some words which have both a technical and a general meaning.

There are five questions for you to answer. Write your answers clearly and unambiguously so that your reader knows exactly what you think. Remember that the questions are asking what **you** think; they are not asking for the 'correct' answer; (in a sense there is **no** single correct answer, since the meaning varies with the context). Later in the course, when your answers have been handed in, you will be provided with our answers, so that you know how **we** think the words should be used. You should compare your answers with ours, and reflect on any differences.

Be concise. Each question should be answered in a few lines.

Questions

1. An analytical procedure needs to be *reproducible, precise, accurate, sensitive*.
How would you investigate how well a procedure meets these criteria?

2. Explain why *systematic error* is harder to detect than *random error.*
3. Under what circumstances would you describe an error as *negligible*, and a difference between two values as *significant* or *insignificant*?

4. Can a *qualitative* procedure prove that a constituent is absent from a substance? And can a *quantitative* method be used to determine exactly how much of a constituent is present?

5. How would you decide whether data are *correlated* and when would you consider *transforming* data?

Notes
* in order to be *correlated*, data must have an x value and a y value; the question here is: under what circumstances are the two values correlated with each other?
* if you measure something (e.g. temperature), you may sometimes wish to *transform* it (for example to 1/T)

You may use examples to illustrate your answers, if you find this appropriate.

Figure 1

of the course and any handed in before the end of term were accepted for evaluation. We do not think the responses were affected either by the content of the course or by the four-lecture course on Analytical Procedures which coincided with the beginning of the laboratory course, since neither were designed to deal with the kinds of question we asked in the questionnaire. In order to encourage students to provide answers that reflected their current understanding, and to discourage them from seeking textbook answers, we made it clear that the exercise was voluntary and that answers would not contribute towards the mark for their laboratory work.

We did not attempt to analyse the overall response of each individual since our intention was not to attempt to map individual understanding of errors but to gain an overview of the sorts of ideas and misconceptions that students in general have about errors. We were concerned not only to discern how much understanding the students have, but also the nature of any misunderstanding. Accordingly, we evaluated how well each response answered the question and we also looked for answers that demonstrated some understanding of the issues even though the wording was more indicative of declarative knowledge than of procedural understanding. Before attempting to evaluate the student responses, we drew up a table giving a short (one sentence) acceptable answer to questions about each of the thirteen words or concepts. This defined the key points we looked for and helped us to be consistent in our evaluation.

Results

A total of 33 responses were received. This rather low response rate (32%) is almost certainly a consequence of the explicitly voluntary nature of the exercise. However, the average A-level score of the respondents was 22 points (three subjects, excluding General Studies, equivalent to BBC) compared with 24 points (equivalent to BBB) for the whole cohort. Thus, on the basis of the only criterion available, the respondents are a reasonable cross section of the whole cohort.

We classified all the responses as showing either 'some or good' understanding, or 'little or no' understanding. Our attempts to attain greater precision, for example by classifying under four rather than two headings, were unsatisfactory. It therefore seemed better to present a crude summary followed by a more detailed discussion of the student responses to each of the five questions in turn. Figure 2 shows the summary of our findings. Even though we were generous in attributing 'some understanding' to some responses, less than 50% of the respondents were judged to show 'some or good understanding' of most of the terms. In the discussion which follows we first enumerate and describe those responses which show 'some or good understanding', and then those which show 'little or none'.

Summary of responses from 33 students

Word or concept	Good or some understanding	Little or no understanding
Reproducible	18	15
Precise	13	20
Accuracy	14	19
Sensitive	6	27
Random error	13	20
Systematic error	26	7
Negligible	16	17
Significant difference	20	13
Qualitative	6	27
Quantitative	7	26
Correlated	31	2
Transforming data	31	2

Figure 2

Question 1 An analytical procedure needs to be *reproducible, precise, accurate and sensitive*.
How would you investigate how well a procedure meets these criteria?

Specimen answer:

- Reproducible: Make multiple measurements on same sample using same procedure.
- Precise: Determine by observation (of replicate measurements) how many significant figures are justifiable.
- Accurate: Use procedure on a standard sample to check closeness to correct value.
- Sensitive: Use decreasing quantities or concentrations until signal cannot be distinguished from noise.

The question of investigating **reproducibility** was significantly better answered than the other three concepts; eighteen of the respondents based their answers on **replicate measurements,** thus showing 'good understanding'. Five appeared to have confused the reproducibility of results with the opportunity to repeat an experiment (obviously a prerequisite for determining reproducibility, and one which must surely be understood to apply to any analytical procedure). A typical example of this misconception is "*For an experiment to be reproducible it should be easy and affordable to recreate the experiment and experimental conditions*". The remaining ten students (30% of the sample) are judged to have no useful understanding of any of the four terms dealt with in question 1 because they either made no attempt to differentiate between them or they dealt specifically with reproducibility but did not distinguish between precision, accuracy, and sensitivity. Examples of these responses are "*Repeat the procedure a number of times to see if there is a large error in the accuracy etc. in which case the results would be significantly different*" and "*In order for an experiment to be reproducible all measurable* *external stimuli should be measured and taken into account. Precision, accuracy and sensitivity can be obtained using a suitable instrumentation giving an acceptable degree of accuracy and a method of measurement which reacts quickly enough to observe significant changes."*

In describing how to investigate the **precision** of a measurement, only seven respondents gave some indication that the key indicator is the number of significant figures that can be justified. A lower level of understanding was demonstrated by six students who made some reference to repeating the procedure several times to obtain the precision (a point already made by three of them in connection with reproducibility); the weakness of these answers was that none gave any indication of how they would judge the precision, or indeed that they understood its meaning in the context of analysis. In addition to the ten already identified as showing no understanding, a further nine showed little understanding, and one gave no response to this part of the question. Four of the nine suggested that precision is related to the care taken with experimental procedures; undoubtedly this is in a sense correct, but it is not a response which engenders confidence that these respondents have a clear understanding of the meaning of precision in this context and it may be related to an assumption that variation is the result of their mistakes. [2] Two confused precision and accuracy and three respondents gave answers, which defied any attempt at classification.

Nine responses on **accuracy** showed good understanding by referring to the use of known or standard samples and comparing experimental results with these. Within this group of nine, four were clearly referring to investigating the accuracy of the procedure using some specific standard sample independently of making an experimental measurement (one of these also offered as an alternative the possibility of making the same measurement using a variety of procedures) and five referred to comparing their results with a literature or text book value (thus illustrating that they are thinking only in terms of analytical exercises to which the answer is already known). Five showed a lower level of understanding by making simple statements about 'closeness to the correct or true value' without giving any indication of how the true value might be known. Nine respondents showed two different sorts of misconception. Five referred to the need to take care either with the procedure or with the choice of equipment, but showed no awareness of the need for calibration or standardisation. The other four confused accuracy with precision (one of them explicitly stating that both accuracy and precision are determined *"from the range of values over which repeats lie"*). Ten showed no understanding, as described above.

Six respondents showed that they understood **sensitivity** to be concerned with the ability of an analytical procedure to detect small quantities or low concentrations of the analyte, though (perhaps not surprisingly) none of these referred to the signal-to-noise ratio as a criterion for judging detection limits. The remaining twenty-seven either showed misconceptions (fifteen) or showed no useful understanding (ten referred to above) or failed to provide an answer (two). Eleven of the fifteen responses with identifiable misconceptions were concerned with the ability of a procedure to detect small **differences** or with the effect on the result obtained of **small changes** in conditions. This is a common meaning of sensitivity in everyday language; however our view is that, in the context of chemical analysis, it is the **precision** of a procedure that determines whether small differences can be detected, and that **sensitivity** is properly reserved to refer to the lower limit of detection. Four responses showed neither useful understanding nor any clear misconception (examples are "*ensuring that all likely changes are measured*" and "*if the reactant is not sensitive to a test, the results will be hard to obtain*").

Question 2. Explain why *systematic error* is harder to detect than *random error.*

Specimen answer:

Random error: Easy to detect from variability of replicate measurements.

Systematic error: Difficult to detect unless you have a reason for supposing the result is incorrect.

The majority of students' responses (twenty-six) showed an understanding that, when systematic errors are present, they occur in all measurements; twenty-one of these linked this with the difficulty in detecting systematic error. Six respondents submitted statements that could not be interpreted as showing any understanding of the use of either 'random' or 'systematic' error. One response included no reference to systematic error.

The responses showed a much lower appreciation of the meaning of random error. Thus only five respondents (a mere 15% of the sample) made clear statements about random error causing variation in readings. An example of these responses is *"random errors are generated by the finite precision of measurement. They affect different readings by different amounts"*. A further eight could (with more generosity) be interpreted as showing some understanding of the term. Four of these stated that random error can be removed by repetition as in *"random error can be worked out of the experiment by repeating it several times"*; these statements could be interpreted as showing that the respondents understand that replicate readings vary as a result of random error, and that the mean value is likely to approximate to the value that would be obtained in the absence of random error. The other four were more difficult to interpret, as in *" random error only occurs in one mode on one variable at one time, and its nature **must** change on repetition so its location and magnitude can easily be determined"*. Fourteen made the mistake of assuming that random error occurs in a single or a small number of results and one of these explicitly described random error as "*a human mistake*". An example of this style of response is "*random error will not affect all results so a result due to random error will look out of place"*. As stated above, the remaining six showed no understanding of the term.

Question 3. Under what circumstances would you describe an error as *negligible*, and a difference between two values as *significant* or *insignificant*?

Specimen answer:

Negligible: When the error is so small compared with the value of the measurement in question that it does not affect the final result (enough for you to care about).

Significant difference: When a statistical (mathematical, objective) test shows that there is only a small chance that the difference between two values arose by chance.

Sixteen responses showed understanding by making some comment to the effect that an error is negligible when it does not have a (noticeable) effect on the overall result (this includes two who related this to whether the overall result is *"close to the expected answer"*, thus drawing attention to the view of many students that analysis involves looking for an already known answer). Three of these sixteen specified a percentage error that would qualify as negligible, but most made little attempt to describe how they would make their judgement as in *"if it does not affect the final result too much"*. Eight other responses specified a percentage error that would be regarded as negligible, but gave no indication of why this was negligible and were therefore judged to be too simplistic to qualify as showing useful understanding. Interestingly the estimates of what might be considered negligible varied from *"several orders of magnitude less than the value"* to *"10%"*, with the most common suggestions lying around 1%. One respondent firmly stated that " *no error is negligible and should always be stated*" – a belief with which we have some sympathy and are inclined to applaud, but it is a very inflexible attitude to apply to the real world of experimentation. One suggested that an error is negligible if it only affects a small proportion of the results – reinforcing that some students regard errors as occasional events rather than as an inevitable feature of measurement. Three more used suspiciously similar wording to state that an error is negligible *"when the result is unaffected whether or not it is included*" (presumably thinking, as the previous response indicated, that the 'error' occurred in one measurement out of a number). The remaining four were so confused as to defy analysis.

In this question we linked the concepts of 'negligible' and 'significance' with the intention of drawing out the point that the latter is almost always concerned with a **difference** between two values (normally mean values), whereas the former (at least in this context) applies to the error in an **individual value**. In practice twenty respondents addressed the question of significant difference in a meaningful way but only five of these showed real understanding of this concept, as in the statement *"when a difference between two values can be explained by error, then it can be regarded as insignificant"*. Fourteen of them stated that a difference would be significant (or insignificant) if it was greater (or less) than a specified percentage of (one of) the values in question. The percentages suggested as a measure of a **significant difference** varied from 1% to 5% (except for one student who did not give a general rule of thumb, but gave as an example that *"the difference between 11032.06 and 11032.91 is insignificant, but that between 1.123 and 1.921 is significant"*). One of them simply stated that a difference is insignificant *"when it won't affect the results"*; we find it hard to assess the level of understanding that this represents. The weakness of all twenty of these answers, which we classified as showing some understanding, is that **none** of them showed any awareness that the amount of variation in (or precision of) data is crucial in deciding whether a difference is significant. Eleven respondents demonstrated that they had not grasped the key point by failing to refer to a difference, and eight of these specifically referred either to a significant **value** or to a significant **error**. One made no attempt to answer this part of the question, and another stated (as an example) that the arithmetic difference between an atomic number and a molar mass is insignificant *"because it has no meaning"* thus demonstrating at least some concern for sensible handling of units!

Question 4. Can a *qualitative* procedure prove that a constituent is absent from a substance? And can a *quantitative* method be used to determine exactly how much of a constituent is present?

Specimen answer:

Qualitative procedure: Can only prove that something is below a certain level (determined by the sensitivity of the method).

Quantitative procedure: Can only determine the quantity within the limits determined by the precision of the method.

Only three students recognised the limitation that very small amounts (below the sensitivity of the procedure) cannot be detected by qualitative methods, and that experimental error prevents exact measurement. In addition the first limitation was recognised by four and the by second three. This left twenty-three respondents who replied affirmatively to both questions. It is true that the wording of many responses might be regarded as ambiguous in a court of law; thus one of the respondents did not state explicitly that a qualitative method can **prove** the absence of a substance, and only nine explicitly stated that a quantitative method could determine **exactly** how much is present. However, wording such as *"a qualitative procedure is used to determine **what** is in a substance"*, or *" a quantitative test can be used to determine the amount of a substance"* does not lead us to believe that the respondent was trying to suggest that the determination is in doubt. Indeed we suggest that the reason why more students specified 'proof' for qualitative procedures but not 'exactly (how much)' for quantitative was that their sentence structure did not require them to add the word 'exactly' to their description of a quantitative procedure, whereas it was harder for them to avoid a word such as 'proof' when describing the testing of absence.

Question 5. How would you decide whether data are *correlated* and when would you consider *transforming* data?

Specimen answer:

Correlated data: Examine a graph of x vs. y to see whether there is evidence of a relationship (correlation).

Transforming data: When the transformation converts a non-linear relationship into a linear one.

Twenty respondents indicated that correlations could be detected from a graph of the data, and a further six said that they would be regarded as correlated if two variables showed some kind of relationship (without specifying how they would detect the relationship). Five said that data are correlated when they fit a mathematical relationship. All of these show some understanding of correlation and only two did not answer this part of the question. Of the twenty who referred to graphical representation of the data, only eight actually answered the question as set by stating that their decision would involve inspecting a graph of the relevant data. The other twelve wrote their answer more in the form of a definition or a theoretical description of correlation. One respondent accepted data to be correlated "*when they show a pattern of increase or decrease or **constancy**"* (our emphasis) and four explicitly restricted the meaning of correlation to linear relationships, (though there is evidence that others shared this misunderstanding even though they did not make it explicit). The wording of the responses (especially the unwillingness to deal with the question "*how would you…*") gave the impression that most students were more familiar with the collection of data known to be correlated than with the concept of investigating whether data are correlated or not.

The question of transforming data gave a similar indication of the majority showing some understanding. Thirty-one referred in a wide variety of ways to the wish to show a relationship more clearly. Many made it clear that the main (or only) purpose was to create a linear relationship from a non-linear one. However the range of answers included a number that indicated more confused objectives almost certainly based on misconceptions. For example three respondents seem to believe that transformation can reveal a correlation which does not exist in the raw data (e.g. *"if there is no correlation, consider transforming the data to determine whether there is any correlation there*"); it may be that these students belong to the group who believe that correlation implies a linear relationship, and several other forms of words suggest that this is the case for a number of others. Many students appear to believe that correlated data can only be analysed when the relationship is linear (*" a plot of T vs. P is useless, but a plot of lnT vs. 1/P is useful"*, or *"data should be transformed to give meaningful graphical data"*), and this includes two who explicitly stated their assumption that the accuracy of their results will be improved by working with a linear relationship. As with the first part of this question, students were apparently reluctant to personalise their answers by answering *"when would you consider…"*. Only two respondents gave no answer to the question about deciding to transform data.

Discussion

Our survey is based on a relatively small sample of students. Nevertheless we believe that our sample is sufficiently representative for us to draw useful conclusions. This confidence is based largely on our comparison of the A-level grades of the respondents and the cohort as a whole. For two of the concepts we tested we are also able to compare our data with the conclusions of Davidowitz et al.[9] These authors analysed the reasons given by 135 second year chemical engineering and science students for making repeat measurements. They concluded that 45% of their sample perceive the purpose to be either to identify a recurring (correct) value (20%) or to perfect measuring skills (25%). This is in broad agreement with our finding that students are more inclined to link variation in measurements with 'mistakes' than with random error (in the sense used in quantitative measurement). The same authors found that more than half of their sample of students, when asked to compare two sets of data, regarded the mean value as of much greater significance than the spread. This is also consistent with our finding that none of our sample of first year students made reference to the spread of data when discussing significant differences.

Our data indicate that first year chemistry students would benefit from a considerably better understanding of the language used to deal with error in quantitative measurement. We believe that this is a matter for concern since the handling of quantitative data is of crucial importance to the procedural understanding of science. We suggest that teaching in this area needs to be radically rethought and restructured because the problem is not simply one of impressing correct ideas about errors on to a blank sheet of a student's mind, but to reconstruct their misconceptions into mature understanding. The first step must be a careful analysis of the key concepts, which students need to understand. We do not claim that our list of five questions covers all of these key concepts. For example, with the benefit of hindsight, we recognise the value of directly probing the students' perception of the origin of variation in measurement. However our study illustrates the value of covering qualitative aspects of the use of language (such as 'negligible') as well as rigorously defined terms (such as 'random error'). Furthermore, the wording of our questions indicates the value of giving meanings and understanding in operational terms like "how would you investigate…" and "explain why… ".

It seems unlikely that students will improve their understanding through textbooks, even if they could be motivated to read them, Our pessimism is based on our inability to find guidance to our questions 3 – 5 in most undergraduate texts. The Open University text on this subject[10] provides a rare example of dealing with the inappropriateness of judging the significance of differences between values by reference only to a mean value and of the benefit of preferring a graph to a table of data in order to discern a correlation between variables. Even for our questions 1 and 2 the guidance given is often contradictory. For example Skoog *et al.*[3] define 'precision' as *"the agreement between two or more measurements that have been carried out in exactly the same fashion"* thus suggesting that a procedure capable of determining a value to only two significant figures is very 'precise' because it lacks the precision needed to detect significant variation. In comparison Atkins[11] defines 'precise measurements' as having *"small random error"* making the cardinal mistake of failing to specify that the error must be small compared with the size of the measurement. Hanson *et al.*[12] state that *"the standard deviation is a measure of the precision of the measurement"*. They go on to use as an example a measure of the boiling point of water for which they quote a mean value of 400.00K and a standard deviation of 0.0126. In contrast the Open University[10] maintains that it is *"rather silly"* to quote a standard deviation to so many significant figures, and there certainly seems little justification for quoting a greater number of decimal points for the standard deviation than for the mean .

Given the lack of clarity and consistency in the textbooks, most chemistry students will necessarily rely on their course work for information about errors and their treatment. Meester and Maskill reported that most laboratory manuals for first year chemistry courses included some information about error analysis but concluded that *"although this indicates the great importance attached to it, generally speaking, error analysis was not a central feature of the courses"*.[13] We do not think the situation has changed significantly in the ten years since the survey was conducted. Furthermore, we have no reason to suppose that the sections on errors in the laboratory manuals are any more likely than the text book accounts to lead to effective learning; our scepticism is based on anecdotal evidence suggesting that there is no consensus amongst academics either about the correct usage of words and concepts used to describe uncertainty in data or in the best procedures available for interpreting experimental data. We are thus led to the conclusion that there is a need for much careful thought about the best ways to meet the Benchmark objective relating to data interpretation.[1] We do not believe that lecturing is likely to be effective because the misconceptions we have documented are unlikely to be corrected by an account (however authoritative) of received wisdom; such accounts rarely involve active participation of the students. Even workshop activities often do little more than introduce the students to routine exercises, which do not really engage their minds. The student perceptions need to be actively challenged in such a way that they reconstruct their own understanding.[7]

We suggest that these operational learning outcomes are most likely to be achieved by regularly and persistently challenging students, through their laboratory work, to discuss their data in terms of the desired outcomes. Thus they could be required to give evidence of the random error in their data, to indicate precautions they have taken to avoid systematic error, to comment on the comparability of data collected by different individuals (and whether differences are significant), and so on. Our proposal is that these regular challenges should be fully integrated into the laboratory course; there is little advantage in paying lip service to the idea by requiring a bolt-on statement about error at the end of a laboratory report. The familiarity with the terms and concepts gained through such regular usage is likely to lead the students to revise their own understanding through a deep learning process. We do not doubt that such learning could be usefully reinforced by structured class discussions and interactive workshops or even by well-constructed lectures, but we suggest that the primary route for learning should be frequent and explicit challenges to use the relevant words and concepts.

Acknowledgements
We acknowledge financial support from the Royal Society for a University Research Fellowship to PJD, the LTSN Physical Sciences Centre for a project grant to JLT, and Glaxo Wellcome plc for support for CJG.

References

1 *General guidelines for the academic review of Bachelors Honours Degree Courses in Chemistry*, Quality Assurance Agency, Gloucester, 1998.
2 M.J. Goedhart and A.H. Verdonk, *J.Chem.Ed*., 1991, **68,** 1005.
3 D.A. Skoog, D.M. West and F.J. Holler, *Fundamentals of Analytical Chemistry*, 5th ed., Saunders College Publishing, Fort Worth, 1996.
4 J.M. Hawkins, J. Weston and J.C. Swannell, *The Oxford Study Dictionary*, Oxford University Press, Oxford, 1992.
5 J. Garratt, A. Horn and J. Tomlinson, *U.Chem.Ed.*, 2000, **4,** 54.
6 G.M. Bodner, *J.Chem.Ed.*, 1986, **63,** 873.
7 K.S. Taber, *U.Chem.Ed.*, 2000, **4,** 63.
8 J. Garratt, D. Clow, A. Hodgson and J. Tomlinson, *Chemistry Education Review,* 1999, **14,** 51.
9 B. Davidowitz, F. Lubben and M. Rollnick, *J.Chem.Ed.*, 2001, **78**, 247.
10 The science foundation team, *The handling of experimental data,* Open University Press, Milton Keynes, 1982.
11 P.W. Atkins and J.A. Beran, *General Chemistry*, W H Freeman, New York, 1992.
12 J.R. Hanson, J.I. Hoppe and W.H. Prichard, *A guide to the reporting of practical work and projects in chemistry,* Royal Society of Chemistry, London, 1995.
13 M.A.M. Meester and R. Maskill, *First year practical classes in undergraduate chemistry courses in England and Wales,* Royal Society of Chemistry, London, 1993.

Using questions to promote active learning in lectures

Communication

William Byers

Faculty of Science, University of Ulster, BT37 0QB.
E-mail: W.Byers@ulst.ac.uk

The first key to wisdom is constant questioning...
By doubting we are led to enquiry, and by enquiry we discern the truth.

Peter Abelard (1079-1142)

An attempt has been made to remedy some of the deficiencies of the traditional didactic lecture by enhancing student involvement and learning through the use of focussed questioning within the lecture format. The potential benefits of questioning are considered and the effectiveness of the approach is evaluated through classroom observations, peer observation, an end of module questionnaire and student discussions. Some limitations of the approach are identified and suggestions for future improvements are made. The paper concludes with a brief consideration of the importance of thinking time to the promotion of meaningful learning.

Introduction

30 years ago when I started teaching I believed that I had knowledge to impart and that the better I taught the more my students would learn. When I, like many others,[1] came to realise that what my students were learning was not always what I was trying to teach them, I tried to teach better. What I then found, however, was that the better I taught the better my teaching was rated by students but not, alas, the better they learned. It was only when I encountered *constructivism*[2, 3, 4] and Alex Johnstone's *Information Processing Model of Learning* (Figure 1)[5, 6] that I started to think about the learner and realised that I needed to teach not just better but differently. Knowledge, alas, can't simply be transferred from the teacher to the learner, much though we might wish that it were otherwise, but meaning must be constructed in the mind of the learner.[2] I see an analogy with digestion where even for a cannibal, ingested proteins are not incorporated directly into body structures but rather are broken down before being reassembled into useful biomolecules. Learning involves the linking and interpretation of incoming information with what is already known by an individual. As we all have different stores of knowledge in our long-term memories (Figure 1) we may all interpret incoming information differently.

If we look into the black box between teaching and learning it seems reasonable that where new information can be satisfactorily linked to pre-existing knowledge and interpreted this will be likely to happen. Piaget[2, 7] referred to this process as *assimilation* and under low resolution it is indistinguishable from simple information transfer. However, this will not always be the case, particularly throughout the education process; confusion

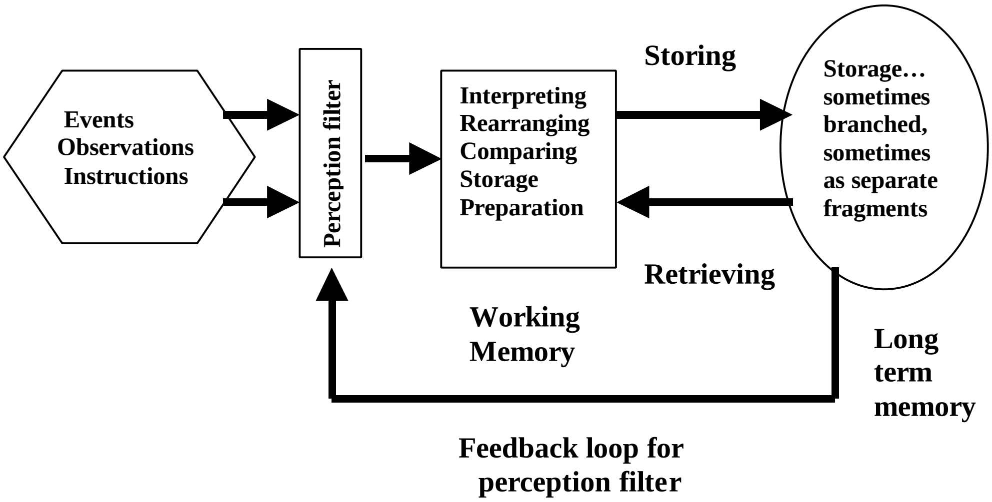

Figure 1. Information processing model for learning

(disequilibrium) will occur whenever new information can't readily be assimilated into existing schemes. What happens now? Such input can simply be ignored, and it is my belief that this is exactly what our students are increasingly choosing to do, it can be linked and interpreted incorrectly leading to so called alterative frameworks[8] or the learner may resort to rote memorisation and try to save the information without linking it to existing knowledge. None of these, of course, represents an acceptable learning outcome from the perspective of the teacher. The possibility also exists however, that the learner will modify the pre-existing schemes until any discrepancies can be resolved. Piaget referred to this process of modifying and developing pre-existing schemes as *accommodation*.[2,7] Education surely involves learning not just more but better and I believe that the promotion and facilitation of accommodation is of pivotal importance in this.[9] The better we organise information and link it to what we already know, the more easily it is likely to be recalled and applied to new situations in the future. However, Herron recently postulated that learning operates on a principle of minimum effort.[10] This suggests that whenever possible the learner will resist restructuring cognitive schemes, preferring to ignore data that don't fit, or to make false connections. Only when the learner becomes really dissatisfied with existing structure will modification start to occur. One can go so far as to suggest that a bad lecture, whatever that may be, followed by peer group discussion or outside reading represents a much richer learning experience than a good lecture, whatever that may be, with minimum student follow-up. Anything which produces active involvement of the learner is therefore likely to enhance the quantity and in particular the quality of learning.

Research by Anderson in 1980 found that questions interspersed in text were amongst the most effective aids to help understanding.[11] I therefore decided to use focussed questioning to try to promote learning within a lecture format. The aim of questioning was not to assess current knowledge or understanding; though undoubtedly some misconceptions, which need to be rectified quickly, will be identified. Rather, the questions sought to promote active learning through the stimulation of thinking and the creation of disequilibrium. Questions can contribute positively in at least four ways:

(a) They may promote both variety in the presentation and more active student involvement during lectures.
(b) They may stimulate learning by relating new information to knowledge already stored in long-term memory (Figure 1).
(c) They may help to identify what is particularly important to concentrate on, *i.e.* they may 'tune the filter' (Figure 1).
(d) They may indicate to the learner where further width or depth of knowledge is needed and help stimulate thought by generating disequilibrium in the learner's mind. Over the years many unsuccessful students have told me that they had expected to do much better in my examinations than they did and that they had understood the material when I covered it in my lectures.[12] They hadn't of course, but like passengers in a car who see no problems with the journey because the driver knows the way, they are incapable of retracing the journey on their own at some future date. It therefore seemed particularly important to break through this complacency.

A question that contributes positively in any of these ways can be considered successful, but I was particularly interested in promoting learning through forcing students to reassess their current knowledge and understanding by creating dissatisfaction with current thinking.

The Study

A 12 week, one hour a week sequence of lectures on bioinorganic chemistry was used to evaluate this approach. The lectures, supported by 12 hours of associated laboratory time, represented one third of a final year honours degree module. The module was taken by 36 full-time students on the BSc (Hons) Applied Biochemical Sciences degree and by 10 part-time students studying for Chemistry or Life Sciences degrees. All students had previously studied some biochemistry as well as chemistry modules, all were well known to me and all but three had been taught by me previously. I started the course with a brief introduction to my ideas about how meaningful learning can take place and to metacognition.[13] The importance of active engagement with new information and participation in class activities was stressed. This occupied about 20 minutes at the start of the first lecture. It was gratifying to note subsequently, that during informal discussion with several students in individual studies advice sessions, students both understood and appeared to support the approach being taken.

Each lecture commenced with a brief synopsis of what was to be covered and one *Big Question* that would be developed and considered during the course of the lecture. The aim of these questions was to prompt students to think about a key problem. Examples of such questions included: *Why are only certain elements used by life?* and *How are ion gradients produced and maintained in the body?* A significant number of questions, about ten each lecture, were also posed during the lectures. These were either targeted at individuals or the class as a whole but were to be answered directly. Research has clearly identified that waiting time must be adequate if questioning in the classroom is to be productive.[14] Care was therefore taken to ensure that adequate time was available and multiple contributions from the class were encouraged. Examples of questions used included: *(1) What is a Lewis acid? (2) Given the solubility product of* $Fe(OH)_3$ *as* 10^{-38} *what is the maximum concentration of free* Fe^{3+} *available in aqueous media at pH7?* and *(3) If copper is the catalytic site what is the likely role of zinc in superoxide dismutase?*

While the first of these questions does appear to require only simple factual recall, its purpose is not to find out that most students on hearing the word acid immediately think of hydrogen ions. Rather it is used to alert the individual that knowledge in long-term memory (Figure 1), which relates to polarisation will be needed. Lewis acidity is a major and recurring theme on the course and related questions used in subsequent lectures included: *Why are metal ions needed when the proton is so effective as a polarising cation? When would 'life' choose to use* Mg^{2+} *rather than* Zn^{2+} *as a Lewis acid catalyst?* and *Why do you think 'life' chooses* Zn^{2+} *rather than* Cu^{2+} *as a Lewis acid catalyst?*

Question 2 is more complex, requiring both conceptual understanding and a 'back of the envelope' calculation. Given time most students should have been able to solve this problem, but in the lecture context progress was only made after some prompting on how best to proceed. In many ways the final question might be considered the most demanding of the three. Here some students had little difficulty in deciding that the most likely role would be structural. Although this question predated any detailed discussion of the biological roles of zinc, a general consideration of roles undertaken by metal ions had previously taken place.

In their own way each of these three very different questions can be thought to show that what is being presented is more complex than it might at first appear and hence create a learning opportunity.

Four approaches were used to evaluate the effectiveness of the approach just described:

(a) Classroom observations.
(b) Peer observation in week 10.
(c) An end of module questionnaire.
(d) A group discussion with three students (four were invited but one was unable to attend) conducted in week 10.

Results

(a) *Classroom Observations*

Having explained the importance of active participation at the onset, I had hoped that willingness to contribute would improve rapidly once students became familiar with the process. However, while five or six students were regularly prepared to contribute, the vast majority avoided answering unless directly challenged. The lullaby effect[15] was apparent with many answers being shallow and not indicative of deep thinking. For example, to the question "*Why is lead not an essential element?*" the answer "*because it is toxic*" seemed to be accepted by all the students. Only after I pointed out that oxygen had been extremely toxic to the earliest forms of life did the class appear to be willing to refocus on cause and effect. Although the level of student contributions fell below what I had hoped for, perhaps it was unreasonable to expect to obtain evidence of meaningful learning concurrent with the teaching (*vide infra*). As an optimist, I can still hope that the stimulus/disequilibrium resulting from the questions may still initiate active engagement with the information over a more appropriate time scale.

(b) *Peer Observation*

As it was the process rather than the content that was of interest, I decided against using a fellow chemist and asked the Coordinator of Learning and Teaching for the Faculty of Business and Management if she would sit in and observe one of my teaching sessions. We met some 15 minutes prior to the lecture and I briefed her on what I was trying to achieve in the session; she made a series of notes throughout the lecture and we met up for a debriefing session later on the same day. Although much good practice was identified, the key observation as far as I was concerned was that I eventually answered all the questions myself. The students knew that I was going to do this and were happy to wait for my answers.

(c) *Student Questionnaire*

The questionnaire asked students to assess the helpfulness of six aspects of the teaching on a six-point scale and then invited free responses relating to the best aspects of the teaching, the worst aspects of the teaching and any suggestions as to how teaching might be improved. The questionnaire was handed out at the end of the last lecture. One student was asked to collect and return the completed questionnaires to me; 34 were subsequently returned. The questionnaire and responses to the six Likert-scale questions are shown in the Appendix.

All six aspects of the teaching, which were evaluated, appeared to be well supported by the class with a significant majority assigning one of the two top grades for each feature. The course booklet, which contained gaps (many of which related to the questioning) to be filled in during the lecture, received outstanding ratings from students returning the questionnaire. They were familiar with the use of structured incomplete handouts as I use this approach throughout my teaching.[16]

Although no textbook was recommended, students were informed that any modern inorganic text was likely to contain a relevant chapter well worth reading. 16 references to original papers were provided and the lecture to which each related, was indicated. Five complete compilations of the 16 references were made available to the class to be used and shared throughout the semester. Several students chose to make their own copies. My discussion on how learning takes place was also generally well received and, as noted above, discussion with students led me to believe that they both understood and supported the ideas outlined.

The usefulness of both the *Big Question* and *frequent questioning* appeared to be highly rated. Hopefully this suggests that even where students were reluctant to voice their opinions they were still thinking about the questions. The free response questions (7, 8 and 9) did provide some interesting information. All respondents identified a best aspect of the teaching. A clear majority (19 students) identified the course booklet, which is primarily a simple information transfer technique, as the best aspect of the teaching. Only four students out of 34 returned questionnaires opted for the questioning. Few students listed any worst aspects though two suggested that there was a lot of material to cover and this led to things being rushed. A further two suggested that starting the lecture at 9.15 on Monday mornings was the worst aspect. Only the following suggestions for improvement were made:

(i) More marks for coursework.

(ii) Review answers to past examination papers to enable students to know what will be required.

(iii) Supplement references with appropriate web page addresses.

While these suggestions all seem reasonable, (i) and (ii) provide further evidence for the assessment-driven motivation for learning which we continue to encounter[17] and (iii) would be more justified if there was evidence that the 16 references provided had been well studied.

(d) *Group Discussion*

A number of issues, mirroring the questionnaire, which I wanted discussed were considered by three of the students. I was there introducing the topics but I took no part in the discussion. The students talked about each issue for some minutes and one student wrote a summary of the consensus view. The group considered that the discussion on learning was a good way to start the course and was useful because it prompted students to think about how they learned. The students thought that the introduction of frequent questions during lectures was beneficial because it helped them to realise how well, or how little, they understood the topics being discussed. The *Big Questions* were also considered useful because they helped to unify each lecture. It was, however, suggested that more active discussion of these questions would help.

Reflective Discussion

Though the approach appeared to attract widespread student support, it clearly did not produce the increased levels of student participation that I had hoped for. It is tempting to suggest that the assessment-driven motivation which directs the behaviour of most students probably means that they did not want to answer my questions, they merely wanted to know my answers. However, perhaps on reflection, my aims were rather ambitious. A majority of students appear likely, initially, to resist any innovative approach to teaching,[18] so an attempt to introduce the questioning approach was unlikely to meet instant success. Clearly, however, if progress is to be made students must be coerced into contributing more effort towards developing their own answers and hence, enhancing their knowledge creation. Unfortunately it seems likely, as recently suggested by Bahor *et. al.*[19] that learning does not occur simultaneously with but after the teaching. This suggests that more success might be encountered if students were required to answer the questions at some time in the future. I have in fact tried to finish lectures with a question, which the class will be required to answer at the start of the next lecture but it was evident that only a few students thought about these questions in the meantime. I believe that I have had success with the use of buzz groups[12] but this is a very time demanding process and thus has to be used sparingly. A recent paper by Hutchinson[20] suggests that awarding some marks for participation will encourage interaction. However, I suspect that Hutchinson's success can probably be attributed to the fact that their students are required to study appropriate chapters before coming to the classes. As this approach was common throughout the general chemistry teaching programme, the students appeared comfortable with the requirements and complied with them. My use of questioning, however, differed from what students encountered in other lectures and more familiarity with the approach is probably needed before progress could be expected. Any attempt to promote more interactive learning will not be straightforward and, unless innovations are introduced with care, may even be counterproductive. I recently heard of a Management module in the third year of an engineering degree where the lecturer asked students to prepare information for oral presentation to the next class. The next class was attended by only 11 of the 110 enrolled students.

The use of student questionnaires for both teaching quality assessments and the evaluation of teaching innovations is now widespread. The present study supports my own experiences over many years that these questionnaires need to be interpreted with caution, particularly the quantitative aspects. I believe that many unwarranted conclusions continue to be drawn from the indiscriminate use of such data.

Peer observation was employed as an assessment tool almost as an afterthought, yet this clearly provided an extremely useful insight into what was actually happening. Perhaps I should have seen what was happening, perhaps I eventually would have; it was certainly clear once the suggestion was made. The experience certainly convinced me that we could all benefit from sympathetic and constructive peer observation and support.

Throughout the course there was constant conflict between time required for questioning and the demands of the curriculum. I can only agree with the two students who suggested that the course was rushed in parts. It seems certain that increased student interaction will require more time than simple didactic teaching. It is therefore perhaps instructive to consider the importance of time to learning.

Some Thoughts on the Relevance of Time to the Learning Process

It is clear from earlier comments that meaningful learning requires effort and is therefore likely also to require time. However, the information-processing model (Figure 1) which has been so useful to understanding how we learn, is, at least in the way that I have viewed it, a time independent model. Yet time is clearly a key variable in determining the quality of learning that can take place. It is surely much easier and therefore quicker merely to transfer information to the learner than for meaningful learning to occur. In fact I suggest that getting the information into the mind of the learner is really the first step towards meaningful learning. This corresponds to what is usually called rote memorising or surface learning. Time and effort are then required to link, interpret, possibly correct (if initially misconstrued) and then accommodate this new information to produce deep or meaningful learning. So, far from being alternatives, rote memorising and meaningful learning may be considered as different stages within the learning process (Figure 2). The second step requires both effort from the learner and time for meaningful learning to develop. The model is clearly consistent with recent suggestions that teaching less, *i.e.* reducing the rate of information transfer, can actually lead to more learning taking place.[21, 22, 23, 24] The model thus predicts that where new concepts are being taught, sufficient time as well as effort is required to enhance cognitive schemes through accommodation, and therefore questions the pedagogical soundness of the recent move towards wide-spread semesterisation in UK universities. Students all too often treat each module as an isolated unit and do not have, or do not take, the time to reflect on what they memorised. It would indeed be tragic if current benchmarking exercises served to increase curricula rather than to embrace and scaffold learning outcomes.

Acknowledgements

The author wishes to thank Kate Greenan for her supportive and insightful comments during the peer observation process, John Garratt for his constructive and encouraging comments on the original manuscript and David Ruddick for many informative discussions.

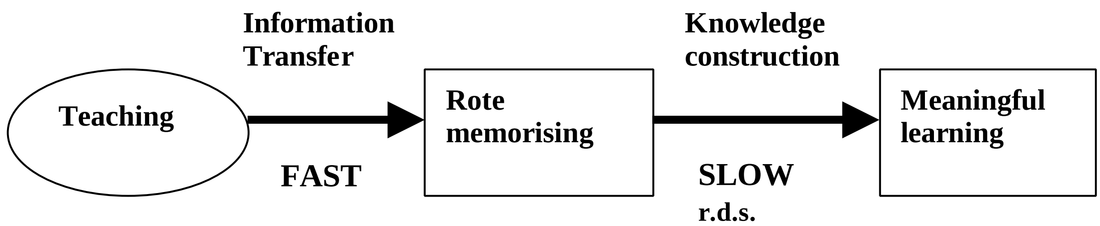

Figure 2 . Model showing the time dependence of learning

Appendix

Bioinorganic Chemistry

Student Opinions on Teaching Approach

Please indicate by ticking appropriate box how helpful you have found each of the following features of the teaching *(stating with 0 to indicate useless rising to 5 where you would consider the feature indispensable).*

	0	**1**	**2**	**3**	**4**	**5**
1. Course booklet					9	25
2. Recommended references		1	5	10	11	7
3. Discussion on how learning takes place		2	3	10	13	4
4. Big question		1	1	13	13	6
5. Frequent questions during lectures			1	10	15	8
6. Prelab session		4	4	6	13	3

Number of responses for each option were as shown above

7. Please indicate what you consider to be the best aspects of the teaching.

8. Please indicate what you consider to be the worst aspects of the teaching.

9. Please indicate how you believe the teaching could be improved for you.

References

1 C.J. Garratt, *U.Chem.Ed.*,1997, **1,** 19
2 G.M. Bodner, *J.Chem.Ed.*, 1986, **63,** 873.
3 L.B. Resnick, *Science*, 1983, **220,** 477.
4 .N. Spencer, *J.Chem.Ed.*, 1999, **76,** 566.
5 A.H. Johnstone, *J.Chem.Ed.*, 1997, **74** 262.
6 A.H. Johnstone, *Educ.Chem.*, 1999, **36**, 45.
7 R. Good, E.K. Mellon and R.A.Kromhout, *J.Chem.Ed.*, 1978, **55,** 688.
8 K.S. Taber, *Educ.Chem.*, 1999, **36**, 135.
9 G.J.Posner, K.A. Strike, P.W. Hewson and W.A. Gertzog, *Science Education* 1982, **66,** 211.
10 D. Herron, *The Chemistry Classroom: Formulas for successful teaching*, American Chemical Society, Washington, DC, 1996, p.18.
11 D. Herron, *The Chemistry Classroom: Formulas for successful teaching*, American Chemical Society Washington, DC, 1996, p.21.
12 W. Byers, Don't Throw the Baby out with the Bathwater, *Proceedings of the 4th European Conference on Research in Chemical Education*, Univ. of York, 1997 p.52.
13 D. Rickey, and A.M. Stacy, *J.Chem.Ed.*, 2000, **77,** 915.
14 M.B. Rowe, *J.Res.Sci.Teach.*, 1974, **11,** 81.
15 M.L. Bigge and S. Shermis, *Learning Theories for Teachers, 6th Ed.*, Longman, New York, London, 1999, p.285.
16 W. Byers, *Proceedings of Variety in Chemistry Teaching 1995*, ed.-s J. Garratt and T. Overton, Royal Society of Chemistry, London.
17 A.H. Johnstone, *Educ.Chem.*, 2000, **37,** 12.
18 R.F. Orzechowski, *J.Col.Sci.Teach.*, 1995, 347.
19 M. Bahar, A.H. Johnstone and R.G. Sutcliffe, *J.Biol.Ed.*, 1999, **33,** 134.
20 J.S. Hutchinson, *U.Chem.Ed.*, 2000, **4,** 1.
21 F. Garafalo and V. Lo Presti, *J.Chem.Ed.*, 1993, **70,** 353.
22 C.J. Garratt, *U.Chem.Ed.*, 1998, **2,** 29.
23 A.B. Arons, *J.Col.Sci.Teach.*, 1980, **10,** 81.
24 A.H. Johnstone and W.Y.Su, *Educ.Chem.*, 1994, **31**, 75.

Why Lecture Demonstrations Are 'Exocharmic' For Both Students And Their Instructors

Perspective

George M. Bodner,

Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA
E-mail: gmbodner@purdue.edu

A theoretical model is proposed to explain why lecture demonstrations are often popular among both students and their instructors. This model provides hints about selecting demonstrations that are most likely to enhance the learning of chemistry. It also suggests ways in which demonstrations can be used more effectively.

Introduction

There are many reasons for doing lecture demonstrations.

- They are fun to do.
- Students like them.
- They grab the students' attention.
- They provide breaks that help students recover from the deluge of information in a typical class.
- They provide concrete examples of abstract concepts.
- Most importantly, they can teach chemistry.

Demonstrations are so attractive they are sometimes done under conditions where neither the students nor the instructors are adequately protected against injury. In an earlier paper we collected examples of accidents and near accidents that might remind chemists of the need to pay more attention to safety when doing demonstrations, hopefully without frightening them away from demonstrations. [1]

Some demonstrations are so much fun for both the students and their instructors that the term *exocharmic* has been used to describe these demonstrations that are so inherently fascinating they "exude charm".[2] The thermite reaction might be an example of an exocharmic demonstration.

$$Fe_2O_3(s) + 2\ Al(s)\ ?\ \ Al_2O_3(s) + 2\ Fe(l)$$

In a lecture manual developed for use at Purdue University, we describe various ways in which this popular demonstration can be used.[3] We argue that it can be used as the basis for discussions of the chemistry of the elements; to demonstrate what we mean by the term *exothermic*; to convince students that aluminum is not 'inert', regardless of their experience with sandwiches wrapped in aluminium foil; or to help students develop an appreciation of what we mean when say that a reaction gives off approximately 800 kJ/mol.

This paper is based on the assumption that none of these uses satisfactorily explains the enormous attraction that demonstrations of the thermite reaction have for both students and their instructors. Furthermore, it assumes that it might be useful to understand the fascination of 'exocharmic' demonstrations such as th e thermite reaction, so that demonstrations can be used more effectively to teach chemistry. It therefore proposes a model based on a theory of motivation, which assumes that these demonstrations fall into the category of phenomena known as *discrepant events*.

A Theory Of Motivation

When I was a student, the most hated words in the English language were 'intuitively obvious' because they were invariably used to describe things that were never obvious to me. When I became a teacher, the most hated words became: "Is this going to be on the exam?" Often, but not always, students' use of this phrase stems from their questioning the value of material we ask them to learn because they don 't think it is important. We 've seen this behavior with students of all ages, from elementary school through the final stages of graduate work. It is frequently a sign of the instructor's failure to motivate the students to *want to learn*.

Motivation is a complex topic.[4] One aspect of motivation, however, can be understood in terms of the theory of optimal arousal.[5] This theory assumes that we try to attain a state in which we experience *some* arousal of our senses — not too little, nor too much. At times, we devote considerable effort to raising our arousal level by reading exciting stories; by going to scary movies; by riding roller-coasters; and so on. Other times, we escape situations where too much is happening to seek peace and quiet. It isn't just the frequency and intensity of the input our senses receive that determines arousal. Th e *content* of this sensory input is also important. Consider what would happen if you put down that fascinating document on household contents insurance and picked up a novel you found interesting. There would be no change in the frequency or intensity of the input your senses would receive. (No more light would reach your eyes, for example.) But it is likely that there would be a change in your level of arousal.

What makes us respond to an object or event isn't the physical input of our senses as such, but a difference between what we experience and what we expect. In other words, we respond to situations that have an element of *surprise*.

We tend to like mild surprises, not severe ones. If there is no surprise, there is too little arousal and we feel bored. If there is too much surprise, we feel shocked and disoriented. With due apologies to the author of the story of Goldilocks and the Three Bears, this theory assumes that we avoid extremes of either too much or too little surprise, and tend toward an intermediate stage in which the amount of surprise is 'just right.'

There is abundant evidence that we become habituated to events that occur in a regular schedule to the point that we ignore them. (We no longer respond when the event occurs.) One of the most common occurrences of this phenomenon in the classroom involves rhetorical questions. It has been shown that many teachers fail to give students enough time to develop answers to questions they ask.[6] The students soon become habituated to the teacher's tendency to ask questions for which answers aren't expected — rhetorical questions — and from that point on, they don't even notice that questions are asked.

Educators have long recognized the role of the unexpected in motivating students.[7] Curiosity, for example, has been shown to be an important component of learning, particularly among children.[8] But what is curiosity if not a drive to investigate and understand situations that evoke surprise? Individuals in all age groups show a marked preference for objects or situations that are novel; that have an element of surprise or incongruity; that generate uncertainty. One of the simplest ways of introducing surprise into the classroom — and therefore take advantage of students' natural curiosity — is through a phenomenon known as a *discrepant event*.

Discrepant Events

Discrepant events have two characteristic properties: They are contrary to what we intuitively expect and they are events we experience for ourselves. Being told something that is counterintuitive doesn't constitute a discrepant event because we can resolve the conflict between what we hear and what we expect by questioning the validity of what we are told. This is harder to do when we observe the event ourselves.

The thermite reaction can be a discrepant event. Students know that chemical reactions give off energy. (They might even know how to calculate the amount of energy liberated.) But the magnitude of the energy given off in this demonstration and the speed with which it is liberated are counterintuitive. Even those of us who should know better are still surprised by the vigor with which two seemingly 'inert' solids react.

Young children are often surprised by iodine clock reactions[9] when they first encounter them because of the speed with which the solution turns from colorless to deep blue. The 'Old Nassau' demonstration[9] — in which the solution first turns orange and then black — is even more surprising because students don't expect the contents of a beaker to change color twice. Oscillating clock reactions,[10] however, are better examples of discrepant events. Regardless of the extent to which students have been exposed to the concepts of reversible reactions, equilibria, kinetics, thermodynamics and so on, there is nothing in their prior experience that prepares them for a reaction that cycles between states in which the solution is colorless, then gold, and then blue.

Implications Of This Model Of Exocharmic Demonstrations

The notion that some of the fascination of lecture demonstrations results from the fact that they may be discrepant events has an important implication: Demonstrations don't have to be spectacular to be effective. They should, however, contain an element of the unexpected.

Let me offer an example, from my own experience. When I took chemistry for the first time, I was told that equal volumes of different gases contained the same number of particles. Until I took physics, this was the most absurd thing I had heard a teacher claim to be true. I knew that gases contained empty space, but I seriously underestimated the fraction of the space that is empty. It therefore seemed reasonable to expect that equal numbers of gas particles of different size would occupy different amounts of space. I now know I was in good company; Dalton rejected Gay-Lussac's data on combining volumes for the same reason. To me, and to many of my contemporaries, Avogadro's hypothesis was just as counterintuitive as it was to John Dalton.

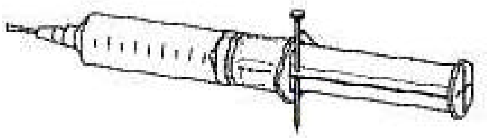

Figure 1

I am reasonably confident that I could have stated Avogadro's hypothesis, if asked to do so on an exam. I am equally confident that I couldn't have used Avogadro's hypothesis to solve a problem because I didn't really believe it to be true.

About 15 years ago, I learned a lecture demonstration that provides a discrepant event that confronts the intuitive model of gases I brought to my first chemistry course.[11] Start with a plastic 50-mL Leur-lok syringe, a syringe cap, and a 10-penny nail. Pull the plunger out of the barrel until the volume reads 50 mL. Now drill a small hole through one of the veins of the plunger into which the nail can be inserted, as shown in Figure 1.

gas	Weight of 50 ml of gas	Number of particles In 50 ml of gas
H_2	0.005 g	1×10^{21}
N_2	0.055 g	1.2×10^{21}
O_2	0.061 g	1.2×10^{21}
Ar	0.081 g	1.2×10^{21}
CO_2	0.088 g	1.2×10^{21}
C_4H_{10}	0.111 g	1.15×10^{21}
CCl_2F_2	0.228 g	1.14×10^{21}

Table 1

Push in the plunger until no gas remains in the syringe, seal the syringe with a syringe cap, pull the plunger back out of the barrel of the syringe, insert the nail into the hole, and weigh the 'empty' syringe to the nearest 0.001 grams with an analytical balance. Fill the syringe with different gases * and determine the weight of 50 mL of each gas. Now use the molar mass of each gas to calculate the number of gas particles in each sample.

Typical data obtained with this apparatus are given in Table 1. Within experimental error, the number of gas particles in each sample is the same. It might still seem strange that equal volumes of different gases contain the same number of particles, but it is no longer possible to avoid this conclusion. Although this demonstration isn't as spectacular as the thermite reaction, or one of the oscillating clocks, it can still be exocharmic because it contains an element of surprise for many students.

Some demonstrations, such as the hydrogen whistle,[12] are such excellent sources of surprise that

* A simple way to handle gases for lecture demonstrations starts with a 1 or 2 inch length of ¾-inch diameter plastic tube. Plug one end of the tube with a rubber septum and secure the septum to the tube with copper wire. Fill a balloon with the appropriate gas from a gas cylinder and insert the open end of the plastic tube into the mouth of the balloon. A sample of the gas can now be collected with a syringe by inserting the syringe needle through the rubber septum

nothing has to be done to enhance their status as discrepant events. The hydrogen whistle is based on the apparatus in Figure 2, which consists of a pair of metal funnels welded together so that there is a small hole at the top and a somewhat larger hole in the bottom. The apparatus is filled with H_2, the hole at the top is plugged with a match, and a rubber stopper is used to close the hole at the bottom. The lights are then dimmed, the stopper and match are removed, and the match is used to ignite the H_2 that escapes through the hole at the top. Attention is drawn to the small flame at the top of the apparatus and the students are told to listen carefully. As the H_2 is consumed, air rushing in through the hole in the bottom makes the apparatus vibrate, and a clear 'whistle' can be heard.

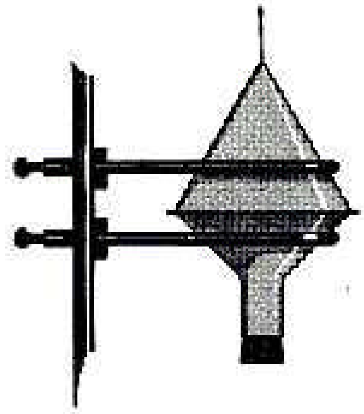

Figure 2

The frequency of the whistle changes with time, as the average molecular weight of the gas in the apparatus increases. With the lights dimmed, and the students paying careful attention to the change in the frequency of the low-intensity whistle the apparatus emits, the demonstration becomes 'striking' when the gas in the container reaches one of the explosive mixtures of H_2 and O_2 — and a loud detonation is heard as a flame shoots out of the bottom of the apparatus.

Demonstrations And The Theory Of Cognitive Change

Some might argue that demonstrations, by themselves, are sufficiently powerful as a teaching device that all we have to worry about is simply doing them. Sarason, however, suggested the following rule for curriculum development or curriculum reform: "A good idea, whose time has come, is no guarantee of success".[13] We'd like to propose a corollary to Sarason's rule: An ideal demonstration, done 'properly', is no guarantee that students will learn what we thought we demonstrated.

Instead of having the students play the role of passive observers of a demonstration, we might use the demonstration as the basis of a phenomenon that White and Gunstone describe as a POE task — from *Prediction*, *Observation*, and *Explanation*.[14] The first step in a POE task is to ask students to predict the outcome of some event, such as what might happen during a demonstration. They are then asked to describe what they observe, and finally asked to reconcile any conflict between what they predict and what they observe. Much has been written in recent years about the misconceptions students bring to chemistry[15] and the fact that these misconceptions are difficult to change.[16] Demonstrations, by themselves, won't overcome misconceptions, but they can provide the basis on which conceptual change is built.

Strike and Posner[17] have proposed a model of conceptual change that begins when students become *dissatisfied* with their present concept. They argue that dissatisfaction is necessary for conceptual change to occur, but not sufficient to induce the change. The student must *understand* the new concept they have been asked to learn. The new concept must also seem *plausible* to the students. And, the new concept must seem *fruitful* — it must seem worth learning. The demonstration of Avogadro's hypothesis in Figure 1 provides the basis for the first step in this model, the stage at which students begin to question the conceptual understanding they bring to the course.

Conclusion

If you accept the arguments in this paper, you can think about demonstrations in terms of the following guidelines.

- There is no evidence that students learn from demonstrations, by themselves.
- There is some evidence that students remember the visual images of a demonstration long after

they forget the words.

- Good demonstrations provide a basis on which learning can be built.
- Demonstrations don't have to be spectacular — or dangerous — to be useful.
- Demonstrations that contain an element of surprise, which don't behave the way students might expect, are often the most charming.
- Demonstrations that are the most charming are those that students remember.
- Demonstrations that are charming can therefore facilitate both the learning of chemistry and the retention of this knowledge.
- Demonstrations that students find 'exocharmic' might therefore be those that best teach chemistry.

You might also note that the spectacular (and often dangerous!) demonstrations that attract some students to chemistry might be driving others away by giving students an unrealistic image of what chemists do.

References

1 G.M. Bodner, *J.Chem.Ed.*, 1985, **62**, 1105.
2 R.W. Ramette, *J.Chem.Ed.*, 1980, **57**, 68.
3 G.M. Bodner, P.E. Smith, K.L. Keyes and T.J. Greenbowe, *The Purdue Lecture Demonstration Handbook, 6th Edition.* West Lafayette, IN, 1997.
4 R.J. Ward and G.M. Bodner, *J.Chem.Ed.*, 1993, **70**, 198.
5 D.G. Mook, *Motivation: The Organization of Action*, W. W. Norton, New York, 1987.
6 M.B. Rowe, *Teaching Science as Continuous Inquiry*, McGraw-Hill, New York, 1978.
7 R.L. Shrigley, *Sci. and Children*, 1984, **21**(7), 6.
8 J.J. Koran and S.J. Logino, *Sci. and Children*, 1982, **20**(2), 18.
9 H.N. Alyea and F.B. Dutton, *Tested Demonstrations in Chemistry*. Journal of Chemical Education, Easton, PA, 1965, p. 19.
10 B.Z. Shakhashiri, *Chemical Demonstrations: A Handbook for Teachers of Chemistry*. Volume 2. University of Wisconsin Press, Madison, WI, 1985, pp. 232-307.
11 I. Talesnick, 8th Biennial Conference on Chemical Education, Storrs, Connecticut, 1984,
12 R.D. Eddy, *J.Chem.Ed.*, 1959, **36**, 256.
13 S.B. Sarason, *The Predictable Failure of Educational Reform: Can We Change Course Before It's Too Late?* Jossey-Bass, San Francisco, 1990.
14 R. White and R. Gunstone, *Probing Understanding*, The Falmer Press, London, Chapter 3, 1992.
15 M.B. Nakhleh, *J.Chem.Ed.*,1992, **69,** 191.
16 G.M. Bodner, *J.Chem.Ed.*, 1991, **68,** 85.
17 K.A. Strike and G.J. Posner, A Conceptual Change View of Learning and Understanding, in *Cognitive Structure and Conceptual Change*, L.H.T. West, and A.L. Pines,, Ed., Academic Press, New York, 1985.

Science and the public: Teaching *about* chemistry on university courses

Perspective

Jim Ryder

Centre for Studies in Science and Mathematics Education, School of Education, University of Leeds, Leeds, LS2 9JT, United Kingdom
E-mail: j.ryder@education.leeds.ac.uk

A distinction can be drawn between knowledge *of* chemistry (the facts, concepts and relationships of chemistry, e.g. the structure of benzene, valency, Raoult's law) and knowledge *about* chemistry (the practices of chemistry, e.g. how chemists decide which questions to investigate, how new knowledge claims in chemistry are developed and validated and how disagreements between chemists are resolved). Such knowledge about chemistry is of relevance to all chemistry undergraduate students irrespective of their future employment intentions. Whilst knowledge about chemistry is inevitably an aspect of university chemistry courses, it is suggested that knowledge about chemistry needs to be taught *explicitly* and should be a *recurring* feature of university chemistry teaching.

Science and the public

There is growing interest both in the UK and worldwide in the ways in which science interacts with public policy. The relevance of this issue has been highlighted in recent debates, such as those concerning the safety of foodstuffs derived from genetically modified organisms, whether or not depleted uranium used in warheads might be a cause of leukaemia amongst military personnel and the safety or otherwise of the measles-mumps-rubella (MMR) vaccine. In all these cases appeals have been made to scientists to provide evidence to inform public debate. On the morning I began work on this perspective I had listened to an interviewer on a national radio programme questioning two scientists concerning their work on the toxic and radioactive effects of depleted uranium on humans, an excellent opportunity for contemporary science findings to inform media debate. However, from the listener's point of view, the key feature that emerged from the interviews was that the two scientists disagreed about the conclusions that could be drawn from their work. What is the listener to make of this? Is one of the scientists incompetent or even biased? What the listener, and perhaps also the scientists being interviewed and the interviewer herself, needed was some understanding about how science works, i.e. knowledge *about* science. This needs to be part of people's general understanding about science to enable them to engage in science-related debates as they arise. In the context of the radio interview the listener should be able to appreciate that the question of the impact of depleted uranium on human health is a complex one. Carrying out empirical work in this area using human subjects is not an option. Whilst empirical investigation might involve non-human subjects or *in vitro* studies, such work is open to questions about the validity of extrapolating its results to humans. A retrospective investigation might involve a statistical study of the health of military personnel and relating it to their exposure to depleted uranium. Here issues such as sample size, estimating dosage and the location of exposure in the human body, become important. Also, cases of leukaemia may have occurred as a result of other causes. How can these be distinguished from those that might follow from exposure to depleted uranium? All these considerations involve knowledge about how science works as much as they do technical knowledge of uranium and its physiological effects on humans.

A number of detailed studies[1] have been made of how non-scientists make decisions on issues with a scientific dimension. Examples involving chemistry include local debates about the toxicity of emissions from an industrial site located near to urban housing and the impact and causes of acid rain. As in the depleted uranium case, these studies show that the knowledge important in these issues is not solely, or even significantly, knowledge of the facts of science. It is knowledge *about* science that often plays the most crucial role as scientists, mediators of science, and the public become involved with science as it affects issues of public policy.

Purposes of university chemistry courses

University chemistry courses provide preparation for three broad areas of employment: as professional research chemists; chemistry-related employment in the media, teaching, the commercial sector, or within local or national government; and employment not directly related to chemistry, such as work in business and finance sectors. Where science relates to issues of public interest, such as in the examples given earlier, individuals in all three employment areas may be involved. Research chemists generate new findings and are asked to report on these to their peers, their funders and the public. Science journalists, spokespeople for commercial companies and pressure groups, and science policy makers provide their reactions to the findings of the scientists. Those not professionally involved in science react to the findings by making choices as consumers, protesters and/or voters. In many contexts public response can have a significant impact on the direction of future science research as commercial and governmental organisations react to consumer and voter pressure. In this way all graduate chemists have a role to play in public debates about science. Given the crucial role of knowledg e *about* science in such debates, its incorporation into the curriculum would be a service to all chemistry undergraduates.

Additional impacts of knowledge about chemistry

Aside from the science and public policy rationale outlined above, there are additional, perhaps more immediate, reasons for developing students' ideas about how chemistry works. There is growing evidence that encouraging students to think about the structure and purposes of scientific knowledge can support their understanding of science concepts. For example, one study[2] designed and evaluated an upper secondary course that included teaching about the general relationships between theory and phenomena in science alongside the teaching of energy transfer in electrical circuits. For many students an understanding of the nature of scientific knowledge enhanced their ability to apply their developing understanding of the concept of energy in electrical circuits in experimental situations. To my knowledge, the interaction between knowledge about science and science concept learning has yet to be examined within university science courses. By contrast, the interaction between ideas about science and university science students' experiences of investigative work has been examined. A study involving chemistry undergraduates found that naïve views about how data and theoretical models interact in science can act as a barrier to students' progress during investigative project work.[3] It is likely that emphasising knowledge about chemistry within university courses will enhance students' understanding of chemistry concepts and their actions during investigative work.

Knowledge about chemistry in the curriculum

Associated with continuing concern about the nature of the interaction between science and those not professionally involved in science [4, 5] there have been a number of initiatives to emphasise knowledge about science within pre-university science education.[6] For example, the current National Curriculum for Science in England has a new section entitled 'ideas and evidence in science' that focuses on 'how science works'. At the post-16 level there is a new AS course 'Science for Public Understanding';[7] a group supported by the Royal Society has begun investigations towards an AS course on the 'History and Philosophy of Science'; and a project funded by the Nuffield Foundation has recently published materials for teaching about science *within* A level science courses.[8] Similar projects have been pursued outside the UK.[9] Against this background, new entrants to university chemistry courses will increasingly be aware of discussions about how knowledge in chemistry is developed, how disputes in chemistry arise and are resolved and what chemistry knowledge can and cannot contribute in complex problems outside the laboratory whenever chemistry interacts with issues of public concern. In part this article aims to contribute to a debate about whether/how university chemistry courses should respond to these developments.

University students' knowledge about chemistry

It might be said that 'how chemistry works' is addressed already in university courses. Indeed any course that requires students to apply chemical knowledge in problem solving tasks, and to undertake science investigations of their own, is inevitably raising issues of knowledge about chemistry. However, many of the areas in which students are asked to solve problems or conduct chemical investigations are far removed from the ways in which chemistry impacts on public policy. For example, many first and second year courses involve investigations in the laboratory in which the answer is already known and the detailed guidance notes limit the chances of things going wrong. Of course such activities provide a legitimate way of developing students' ability to use established empirical techniques. However, they are unlikely to communicate the uncertainties and complexities of applying chemistry to issues of public concern.

Furthermore, despite the inevitable presence of knowledge about chemistry in university courses, several studies have shown that students often leave university with very naïve views about how science works.[10,11] Many students see science as capable of providing 'hard facts', that it is always possible to obtain data that will provide a single, incontrovertible interpretation. The presence of uncertainty and multiple interpretations, particularly in complex settings, is often not recognised.

Teaching knowledge about chemistry

So how can knowledge about chemistry be communicated within undergraduate courses? The strongest message coming from the few studies that have been conducted to date is that knowledge about how chemistry works needs to be taught *explicitly*. It is rarely sufficient for students to engage in chemical investigations or chemical problem solving activities for them to develop their knowledge about chemistry. For example, we followed the experiences of 11 undergraduate science students (including 2 chemists) as they undertook final year research projects over a period of 8 months.[12] These projects gave students the experience of engaging in authentic research that addressed complex issues. Gathering reliable data was often a real challenge and in most cases only tentative conclusions could be drawn. However, experiencing authentic research was found not to be a sufficient condition for being able to articulate an appropriate view about 'how science works'. Many of these students persisted with their view of science as always involving 'hard facts'.

To make knowledge about chemistry explicit students need to be encouraged to ask questions about the structure, purpose and limitations of chemistry knowledge. How sure can we be about our conclusions? Do our findings enable us to make any generalisations outside the context of the study? What can our laboratory study tell us about the chemistry of materials in contexts outside the laboratory? What additional issues would need to be considered? Are there other possible interpretations of the data? If so, what should chemists do next in order to resolve the dispute? It might be said that add-on courses on the History and Philosophy of Science/Chemistry could serve this function. I would argue that whilst such courses do serve legitimate aims, they are not best placed to develop students' ideas about science with a view to supporting their consideration of public policy issues and their learning of science concepts and investigative activities. The teaching of knowledge about chemistry needs to be an integrated part of a university chemistry course, with discussions about how chemistry works running through lecture courses, problem solving classes, investigative work, and (critically) assessment activities.

Conclusion

Not everyone would agree that a strengthening of knowledge about chemistry within university courses is desirable, particularly given the other pressures on curriculum time. This perspective aims to contribute to a debate about how/whether university chemistry courses should respond to the increasing focus on knowledge about science within pre-university education. Secondly, lecturers themselves have limited experience in explicit teaching about chemistry and few resources are available to support such teaching. This perspective is also a plea for university chemistry teachers who recognise the need for knowledge about chemistry teaching to develop such teaching and to communicate to others what works and what doesn't work and there are signs that this is beginning to happen.[13] Finally, there has been little, if any, research into the impact of knowledge about chemistry teaching on students' ideas about chemistry within university courses. Such studies would provide insights into how students develop new ways of thinking about the practices of chemistry both within chemistry research and also in matters of broad public concern.

References

1 J. Ryder, *Studies in Science Education*, 2001, **36,** (in the press)

2 A. Tiberghien and O. Megalakaki, *European Journal of Psychology of Education*, 1995, **10,** 369.

3 J. Ryder and J. Leach, *International Journal of Science Education*, 1999, **21,** 945.

4 House of Lords, *Science and society,* Third report of the Select Committee on Science and Technology UK, 2000.

5 Office of Science and Technology / The Wellcome Trust, *Science and the Public: A review of science communication and public attitudes to science in Britain*, 2000.

6 R. Millar and J. Osborne, (eds.), *Beyond 2000: science education for the future,* King's College, London, 1998.

7 A. Hunt and R. Millar (eds.) , *AS Science for Public Understanding*, Heinemann, Oxford, 2000.

8 A. Hind, J. Leach and J. Ryder, *Teaching about science*, A research project funded by The Nuffield Foundation, 2001 (materials are available at: http://www.nuffieldfoundation.org/aboutscience/index.shtml)

9 W. De Vos and J. Reiding, *International Journal of Science Education,* 1999, **21,** 711.

10 N.W. Brickhouse, Z.R. Dagher, W.J. Letts and H.L. Shipman , *Journal of Research in Science Teaching*, 2000, **37,** 340.

11 J. Ryder, J. Leach and R. Driver , *Journal of Research in Science Teaching,* 1999, **36,** 201.

12 J. Ryder and J. Leach, *International Journal of Science Education*, 1999, **21,** 945

13 *E.g.* J. Garratt, T. Overton and T. Threlfall , *A question of chemistry. Creative problems for critical thinkers*, Longman, 1999; J. Garratt, *U.Chem.Ed*., 1997, **1**, 19; S.T. Belt and L.E. Phipps, *U.Chem.Ed*., 1998, **2**, 16; S.B. Duckett, N.D. Lowe and P.C.Taylor, *U.Chem.Ed*., 1998, **2**, 45; S.T. Belt, M.J. Clarke and L.E. Phipps, *U.Chem.Ed*., 1999, **3**, 52.

Customising and Networking Multimedia Resources

Letters

From Antony J Rest
Department of Chemistry,
University of Southampton,
Southampton SO17 1BJ.
a.j.rest@soton.ac.uk
http://www.soton.ac.uk/~ajrchem

The article by George McKelvy, 'Preparing for the chemistry laboratory: an Internet presentation and assessment tool'[1], which describes networking multimedia images and resources at the Georgia Institute of Technology, prompts the question "Is such practice common in the UK and Europe?"

The short answer is 'no'. The UK is, however, at the forefront of such developments through the CTI Centre for Chemistry, the 'Teaching and Learning Technology Programme' (HEFCE TLTP), the 'Funds for the Development of Teaching' (HEFCE FDTL) projects and Government initiatives to provide computers for schools and colleges, as described in the SOCRATES Open and Distance Learning Report.[2] However, even in the UK few universities have attempted schemes as ambitious as that described for Georgia. The prime reasons for this would seem to be lack of time and resources. This raises the question "How can such obstacles be overcome?"

To create the resources the choice is either making one's own or obtaining, customising and networking images and materials. The latter requires the consent of the Copyright owner. For this reason Georgia Institute of Technology, which has very large freshman classes, could afford to make its own materials. This is an expensive option. For example it cost £500,000 to make the video images for the "Basic Laboratory Chemistry" laser video discs and VHS tapes and £40,000 to customise them with interactive materials for the series of CD ROMs: "Practical Laboratory Chemistry" (http://www.emf-v.com). In favour of the other route is that most publishers and copyright holders of images, e.g. the "Chemistry Images" database (http://www.rsc.org/is/cvc/chem_img.htm), are amenable to requests to customise and network materials within an institution, provided that such an institution purchases a copy of the materials, does not 'export' them to other institutions and that a modest licensing agreement is signed. This makes the customising route much cheaper and saves "re-inventing wheels" but the costs cannot be recouped by selling on the materials.

Customising existing materials is not as difficult as it is often imagined, but ensuring quality is at the heart of producing worthwhile resources. For example compression of video images is best achieved if high quality sources are used (Betacam), good quality software and hardware for capture and compression is used (MPEG), and the compression is not severe but tailored to deliver high quality images and sound. Such images, when captured and digitised, can be stored as files on the hard drive of a PC and incorporated into learning, teaching, and training packages, together with text and animations using a number of design and management packages, e.g. Toolbook and Macromedia Director. Some examples of what is possible will be published in 2001. These may involve compiling a set of images for a specific course from a variety of sources onto a CD ROM, adding subtitles for students with learning difficulties, changing the level of content, e.g. the CD ROMs on '*Practical Chemistry for Schools and Colleges*' (published in 2000) which, in turn, were derived from the '*Practical Laboratory Chemistry*' CD ROMs, and adding extra content as in the series '*Physical Chemistry Experiments*'. Adding subtitles and adapting voice-overs can be applied to produce materials in other languages for scientific and language learning purposes, e.g. a French/English version of '*Practical Laboratory Chemistry*' will be published in 2001.

To maximise availability, networking within an institution is possible. This is a process, which depends on the compatibility of the networking software and the software of the multimedia resource to be networked. Such is the variety of combinations that designing a totally generic package is a daunting task. For this reason most CD ROMs are specified as 'For single user, stand-alone computers'. What is possible, however, is that the video component can be delivered as 'streaming video'[3] to any point within the 'firewall' of an institution, using a dedicated server. Having obtained the video on a server, this can be mixed with learning, teaching and training to produce resources specific to the needs of a particular institution. The use of an internal server ensures that the material remains within the 'firewall' and the copyright conditions set by producers and publishers are met.

The way is open, therefore, for expansion of the use of multimedia materials for learning, teaching and training. This will become crucial in the future because students will increasingly be able to download the resources they need to their own PCs from central network servers.

References

1. G. McKelvy, *U.Chem.Ed.*, 2000, **4**, 46.
2. *Multimedia Resources for Chemistry in Europe,* http://mrc.chem.tue.nl
3. J. Smith, *Viewfinder*, 2000, (40), 12, see also http://webhelp.ucs.ed.ac.uk/services/media/bufvc.html.
   W. Garrison, *Viewfinder*, 2000, (41), 13.
   ["Viewfinder" is the magazine of the British Universities Film and Video Council (BUVFC) and is published four times each

year. It is designated by its issue number. See also http://www.bufvc.ac.uk]

On the need to use the Gibbs' phase rule in the treatment of heterogeneous chemical equilibria

From Professor Paolo Mirone
Department of Chemistry, University of Modena and Reggio Emilia, Modena, Italy
E-mail: mirone.paolo@unimo.it

Misconceptions are well known to be serious barriers to effective learning. Unfortunately it is not just students who have misconceptions. In a recent paper[1] Thomas and Schwenz investigated third and fourth year students' conceptions of equilibrium and fundamental thermodynamics by means of interviews concerning various aspects of a heterogeneous equilibrium. They considered a system at constant temperature and pressure (T: not specified; p = 1 atm). The initial system consisted of $CaCO_3$ and CO_2 only and the students were asked to consider its evolution to the equilibrium system comprising $CaCO_3$, CaO and CO_2. Thomas and Schwentz assumed that the volume of the gas phase at equilibrium would be nearly double that of the initial state, as shown in Figure 1 of their paper.

The behaviour of the system is most easily understood from the Gibbs Phase Rule. The equilibrium state has only two independent components (C), in consequence of the equilibrium condition on the chemical potentials:

$$\mu CaCO_3 = \mu CaO + \mu CO_2$$

Application of the phase rule:

$$F = C - P + 2$$

with C = 2 and the number of phases, P = 3, shows that there is only one degree of freedom, F. Thus if the pressure is fixed experimentally (as Thomas and Schwentz do: p = 1 atm) the equilibrium temperature is fixed (by inference) at 898.6 °C. From a purely thermodynamic point of view, this equilibrium does not differ from the equilibrium between a pure solid and its vapour.

Since pressure and temperature are constant, it follows from the above that, provided the values of the physical variables p and T are compatible with equilibrium, then equilibrium is reached as soon as the very first crystals of calcium oxide have formed, namely, before the amount of additional carbon dioxide is enough to show any appreciable volume increase. So, the process is not "*a real spontaneous change*" as stated by Thomas and Schwenz.

It is interesting to consider what would happen if the temperature were above 898.6°C, but we have to define the experimental conditions carefully. Suppose the pressure is imposed on the $CaCO_3$ and CO_2 by a piston. The pressure imposed by the piston is 1 atm. At the higher temperature, the equilibrium pressure of CO_2 is above 1 atm, so that $CaCO_3$ decomposes to generate more CO_2 in an attempt to increase its pressure to the new equilibrium value. However, the piston moves back to maintain a pressure of 1 atm, until all of the $CaCO_3$ is converted into CaO and CO_2. Under these conditions, i.e. with an imposed external pressure, $CaCO_3$ has a fixed temperature of decomposition, (898.6°C, at 1 atm) analogous to the boiling point of a liquid.

Nearly half of the interviewees correctly asserted that 'if the temperature were high enough or the pressure low enough, all the $CaCO_3$ would be consumed'. This statement was classified by Thomas and Schwentz as an 'alternative conception' falling into the category 'Using informal prior knowledge from everyday experience to explain the thermodynamics of chemical phenomena'. In other words, they judged that the students had given the wrong answer.

This example shows that, when dealing with heterogeneous equilibria, the first thing to be done is to apply the phase rule. It is simple, powerful and also aesthetically satisfying. It is a pity that it appears so rarely in papers on the teaching of thermodynamics, especially when it can be a key to understanding the system under consideration.

Reference

1. P.L. Thomas and R.W. Schwenz, *Journal of Research in Science Teaching*, 1998, **35,** 1151.

Note Professors Schwenz and Thomas were offered an opportunity to comment, but declined to take it.

Questionable questions

From John Garratt
Sennotts Farm
Scaynes Hill
West Sussex RH17 7NW

I enjoyed reading Byers' Communication on 'using questions to promote active learning in lectures',[1] and am sorry he was disappointed in the results. I hope that others will be encouraged by his report to try something similar and that he perseveres with his own suggestions for improvements. I have two comments that may help to explain why the students participated less actively than he hoped. One relates to habit, and the other to the need for absolute clarity.

Byers introduced his questions in a sequence of twelve lectures attended by 46 students in the final year of their undergraduate course. Compare this with Hutchinson, who refers to a greater level of participation in classes of up to 300 students.[2] Significantly, his class is held in the first year, and it runs at the rate of three lectures a week for 15 weeks, during which time it is the students' only exposure to chemistry. Thus Hutchinson's students quickly become accustomed to the idea that learning chemistry at university involves active participation. In contrast, Byers' final year students have already discovered that learning chemistry at university involves sitting passively in lectures and swatting up the notes just before the exam. They would be more likely to respond to questions in class if the practice of asking them were introduced in the first term, and systematically adopted by all members of staff.

My second point relates to the need for clarity. Students respond much better to questions when they are confident that they know what the question means and what sort of answer is appropriate. That is why, in our book,[3] most of our questions involve giving reasons for selecting one answer from the several we provide. This strategy gives the students clear and useful signposts and seems to help them to get involved quickly. Byers might find this a useful model to use, at least while the students are getting accustomed to the idea of participating in lectures. Also, he might find that students respond better if they are given a moment to discuss the question and its answer with their neighbours. Whether or not he likes these suggestions, I recommend that, before he asks a question in class, he writes down a model answer in the number of words (perhaps 20 or less) he expects from the students in a lecture and which shows the depth of thinking it is reasonable to expect. If he can't do this, then it is unreasonable to expect the students to provide an answer. Take the question "why is lead not an essential element?"; the answer "because it is toxic" is little more than a restatement of the question, and in that sense indicates shallow thinking (as Byers agrees). In fact the real problem lies with the question since the answer cannot be known, nor can it be investigated by observation or by experiment because it lies in evolutionary history. The best we might be able to do is to suggest chemical reasons why lead is incompatible with life as we know it; the question could easily be rephrased to make it clear that this is the desired response. A different style of answer would be something like "life may have evolved in an environment in which no lead was available, and so there was no opportunity to develop a use for it." But few students are likely to be in a position to adopt that train of thought, given that most of them are ignorant of the fundamental principles of life processes and of evolution. I confess to having fallen many times into the trap of asking questions that are ambiguous or require more thought than can be expected in the middle of a class. I also admit that, when I try to provide written model answers, it often takes me many attempts and usually forces me to rewrite the question. It is a salutary lesson and brings home the value of the exercise. Finally, I would like to stress the need for chemists, when asking questions about life processes, to do so from a position of secure knowledge of biochemistry and of evolution.

References

1. W. Byers, *U.Chem.Ed.*, 2001, **5**, 24.
2. J.S. Hutchinson, *U.Chem.Ed.*, 2000, **4,** 1.
3. J. Garratt, T. Overton and T. Threlfall, *A question of chemistry. Creative problems for critical thinkers,* Longman, 1999