Stuart W Bennett

Assessment in chemistry and the role of examinations

_____

Stuart W Bennett,

Department of Chemistry, Open University, Milton Keynes, MK7 6AA UK
e-mail: s.w.bennett@open.ac.uk

Introduction

As teachers in higher education, we are increasingly aware that assessment is a (if not the) major driver for students in higher education. Students apply the 'assessment' test. If a concept, skill or knowledge chunk is deemed to be assessable in a way that contributes to the ultimate goal (of degree, diploma etc), then a high priority is accorded in the learning strategy of the student. Given this prevailing culture, we can adopt either of two strategies: to change this culture of assessment-driven learning or to use it as an opportunity to improve learning. Another possibility might be to remove assessment entirely but the arguments for assessment are powerful, embodying ideas of guiding student improvement and progression, diagnosis of faults, providing feedback to teachers and informing employers.[1] Over the last decade and more, there have been many laudable and, to a large extent, successful programmes that have introduced context-based learning,[2] problem solving approaches and holistic perspectives. Nevertheless, even with such approaches, assessment remains a major learning driver for many students. So accepting that there is a strong argument for the retention of assessment, and that changing the culture of assessment-motivated learning would be difficult if not impossible to achieve, a critical consideration of the quality of assessment should be a feature of every study programme.

The vital role of learning outcomes in the design of an assessment strategy is recognised both by teachers and government agencies. The United Kingdom Quality Assurance Agency[3] enshrines this link in both recommendation and legislation relating to the programmes of higher education institutions in the UK. The Agency defines learning outcomes as 'statements that predict what learners have gained as a result of learning' and the '…achievement of which a student should be able to demonstrate'. So providing we are able to define learning outcomes competently, students should have a clear idea of what may be assessed and how it is to be assessed. Note that the competent definition of learning outcomes must include information on both the assessment criteria and the mode of assessment. The UK Chemistry Benchmark Statement[4] identifies a range of assessment media. The list includes formal examinations, laboratory reports, problem solving exercises, planning and presentation of oral reports, and the conduct and reporting of individual and

Table 1 Comparison of aspects of formal examinations with non examination-based assessment

Formal examinations	Non-examination-based assessment
Allow for verification of student work	Can be less certain that it is the student's own work
Performed in limited time (time management)	Time is student-limited
All skills and knowledge may be tested at the same time	Skills and knowledge tested over a longer time but wider range of skills tested
Relatively easy to administer	Skills tested more effectively
Good discriminator on certain criteria	Time for reflection
Relatively easy to grade consistently	Synoptic
Disadvantages some students (poor recall, panicky disposition etc.)	Can be more difficult to administer
Unable to test some important skills well (selection, organisation, communication etc.)	Reasonable discrimination (but tendency to low standard deviation)
Tests at a particular point in time, no measure of retention	Difficult to grade consistently
Choice of questions can mean that some areas not tested	Not so memory dependent
Tests memory	Perhaps a more even playing field

collaborative project work, with the possibility of poster displays. There is no shortage of recognised assessment media within the chemistry establishment.

All institutions use a subset of these assessment media and it is important to look at the total learning package in terms of outcomes and assessment. Each medium has its particular strengths in testing specific learning outcomes. One categorisation of assessment media is based on the degree of 'openness' and of 'time constraint'.[5] At one end of the 'spectrum' is the open project with an indefinite (or at least extended) completion date and the other end features the closed book, fixed time, formal examination. Between, there are assessments such as open book examinations, fixed time essays and short projects. However, assessment by formal examinations is distinctive as compared non-examination assessments as can be seen from Table 1.

The following analyses and discussion are directed by three questions:
"Are we teaching what we think we are teaching?"
"Are students learning what we think they are learning?" and
"Are we assessing what we think we are assessing?"
The aim of the study is to determine how well we are using examinations as a measure of our claimed learning outcomes.

Examination and learning outcomes

In this part of our studies, we are focussing on the closed-book, fixed-time examination. Although this is just one of several media, it features in almost all institutions and, with chemistry-based programmes, it makes a significant contribution to overall assessment. Also this assessment medium is generally accessible.

Whilst realising that some learning outcomes are inappropriate for testing by examination, we embarked on a detailed look at first-year university, chemistry-based examination papers and the relationship of questions to learning outcomes. Papers used were from 22 UK universities (58 papers), 6 state universities in the USA (13 papers) and 4 Australian universities (11 papers). The selection of papers from the USA state universities included some early, year-2 module papers, as there is a lower level of subject specialisation at university entrance in the USA than there is in the UK. (An additional 31 examination papers were received but not used, as specific learning outcomes were not specified.)

The first task was to assign the appropriate learning outcomes to the individual questions on the papers. The assignments were carried out in duplicate and the assigning pair of teachers was asked to negotiate over any discrepancies in their question/learning outcome assignment. Very few discrepancies arose (less than 2 per cent) and these were resolved by discussion between the assigning pair, apart from one instance when a third party was brought in. There were many differences in the papers in terms of the numbers of questions, choice of questions and time allocated to questions. The learning outcomes claimed by the institution for the examination and those actually tested by the questions in the examination were tabulated for each paper. The overall learning outcome totals are shown in Table 2.

The following exemplar (Table 3) is typical of an individual paper analysis, representing neither the best correlation between the learning outcomes claimed and those actually tested, nor the worst. The paper comprises a three-hour examination paper with two sections. Students were asked to complete two questions from four from Section A and three questions from five in Section B, a total of five questions to be completed from the nine on the paper. The claim made by module assessment information provided by the institution to the student was that there were seven learning outcomes (which we have designated A-G) tested by the examination.

Table 2 Learning outcomes tested and claimed in examination papers.

Institutions	Number of papers	Total number of questions	Total learning outcomes claimed	Total learning outcomes tested	Outcomes tested/outcomes claimed
4 Australia	11	76	68	39	0.574
22 UK	58	455	179	81	0.452
6 USA	13	96	121	42	0.347

Table 3 Examination questions analysed by learning outcomes.

Learning outcome	Question								
	1	2	3	4	5	6	7	8	9
A	✔	✔		✔					
B			✔			✔		✔	
C		✔			✔		✔		✔
D									
E				✔					
F					✔	✔			✔
G									✔

Learning outcomes A-G
  A   Use given spectroscopic and other data to deduce the structure of molecular species.
  B   Select appropriate reaction sequences to effect specific structural changes in molecules.
  C   Describe all mechanistic steps in functional group transformations.
  D   Interpret kinetic data in terms of reaction mechanisms for some organic transformations.
  E   Assign R, S notation to chiral carbon.
  F   Identify and account for reaction product outcome that is affected by chirality of starting material.
  G   Identify the main features of drug receptor sites explain how selection and specificity is achieved.

This simple analysis indicates that, although there is a claim to test the seven learning outcomes in the examination paper, one outcome, D, is not tested at all. Outcomes E and G are tested in just one question, with the other outcomes being tested in three questions except for outcome C which is covered by four questions. Well, this could be worse as we do have all but one of the outcomes being tested in at least one question.

However, students are not asked to attempt all nine questions. There is a choice. Take a student who attempts Questions 1 and 3 (two from four in section A) and Questions 5, 6 and 8 (three from five in Section B). How does the analysis look now (Table 4)?

A rather less satisfactory pattern now emerges. Although outcomes B and F are tested in three and two questions respectively and A and C are tested in one question each, we now have three learning outcomes, D, E and G, that are not tested at all. This illustrates the need to consider the range of student question selection as well as the totality of the questions actually appearing in the paper. As suggested earlier, the exemplar featured in Tables 3 and 4 is typical of the analysis and certainly by no means is the worst case.

A further major consideration is that of student performance. Threshold pass marks in examinations generally fall in the range 40-45 per cent. With the examination paper structure illustrated, it would be possible for a student to gain a pass with just three of the seven learning outcomes achieved. The point is that we do need to be mindful of what we *claim*. "This examination tests learning outcomes *X, Y, Z* etc." should not be translated as "Students who pass this examination have achieved learning outcomes *X, Y, Z* etc". We might be reluctant to fly if we knew this to be case for the assessment of the training of the pilot of our

Table 4 Examination questions attempted by a student analysed by learning outcomes.

Outcome	Question								
	1	2	3	4	5	6	7	8	9
A	✔	✗		✗					
B			✔			✔		✔	
C		✗			✔		✗		✗
D									
E				✗					
F					✔	✔			✗
G									✗

aeroplane.

The main findings from these simple analyses of examination papers are that:

- There is a mismatch between outcomes claimed and outcomes tested (Table 2 data).
- Some outcomes are tested several times in the same paper and some omitted. This situation is made worse where the paper embodies a choice of questions (Table 3 data).
- In the worst cases, students are able to achieve a pass grade with less than twenty per cent of the learning outcomes fully achieved (Table 4 data).[*]
- Questions that are easy to set and easy to mark tend to predominate (see next section on problem solving in examinations).

Problem solving in examinations

The predominance of the 'easy to set, easy to mark' questions led to a further line of enquiry. These questions tend to be either of the regurgitation of information variety or involve a 'problem' of some sort, often a calculation. This latter type can run from year to year with just a change in the input data. The frequent claim, that the examinations address problem-solving, needs to be looked at in this context.

Many of the calculation-type questions masquerading under the problem solving banner are of the type: calculate the mass of sulfur dioxide produced by burning 1.00 tonne of coal containing 0.700 per cent by mass of sulfur. Certainly this type of question does test a range of skills and knowledge but is it a *problem*? The answer is obtained by applying a simple, standard algorithm: find the mass of sulfur, convert to moles, use in balanced equation to find amount of sulfur dioxide then use relative molecular mass of sulfur dioxide to find the mass of sulfur dioxide. The input data are given, the method is familiar and the output is given.

Problem types have been categorised by Johnstone[6] and others in terms of these parameters and the above can be seen to be a Type 1 'problem'. The 'problems' in Table 5 become more 'problem-like' and less 'exercise-like' further down the table.

---

[*] A few examination papers contained compulsory sections where there was no choice of questions. Others clearly embodied norm-based and criterion-based sections. With these structures, more of the claimed learning outcomes have to be achieved to obtain a pass grade.

Table 5 Categorisation of problem types.

Type	Data	Method	Output
1	G	F	G
2	G	U	G
3	I	F	G
4	I	U	G
5	G	F	O
6	G	U	O
7	I	F	O
8	I	U	O

F = familiar, G = given, I = incomplete, O = open, U = unfamiliar.

An example of a Type 4 problem might be 'How many sugar residues are added per second to a blade of grass as it grows?' or 'How many amino acid residues are added per second to a human hair as it grows?' The inexperienced student, faced with this type of problem might panic but with a little thought it is possible to start confining the broad question. Certainly the input data are incomplete. The method is not immediately familiar but the output is defined by the question. To take the latter problem, how much does hair grow in a second? In a month it probably grows 1-2 cm. People are familiar with roughly how often they have their hair cut. How thick is a human hair? It is certainly less than 1 mm. Would ten hairs side by side cover 1 mm or would it be rather more? So we can get a range for the volume of hair produced in a month (and in one second). What is the density of hair? Probably around 0.8 g cm^{-3} like many organic materials, so we can estimate the mass produced per second. What is the mass of an amino acid residue? Easy, via relative molecular mass, and we thus have the number of residues per second. A bit of thought enables a seemingly impossible problem to broken down into manageable parts. The answer is a staggering number of around 10^{11} per second. So the slow growing hair on the macro scale is a frenzy of activity at the molecular level!

We analysed the questions in all 82 of the examinations papers according to the categories outlined in Table 6. All questions that embodied problem solving (in part or as the whole of the question) were included in the analysis (432 of 627 questions in total), the results featuring in Table 6. Whilst there is an interpretive element into the assignment of questions to problem type, only in very few cases (less than 3 per cent) were there inconsistencies in the allocations recommended by two independent academics and these were resolved by discussion.

Table 6 Categorisation of problems in examination papers.

Type	Number of questions analysed	Proportion/percent
1	409	94.7
2	13	3.01
3	9	2.10
4	0	0.00
5	1	0.20
6	0	0.00
7	0	0.00
8	0	0.00

The claim that 'examinations test problem solving' could be defended if one were to accept that the straightforward, algorithmic type exercise is indeed a 'problem'. However, current interest in problems and problem-solving[7] suggests that the term 'exercise' is more appropriate for type 1 'problems. Our observation that 94.7 per cent of the questions analysed were of Type 1 (data given, method familiar and output given), the standard algorithm type, again suggests a danger in making claims that cannot be substantiated.[*] From this analysis, it does appear that questions that are essentially the same from year to year in which the only change is in the data and not in the structure are very common, easy to set and easy to mark. There are other media that are better suited for the testing of problem-solving but we should not be claiming what we are not doing.

Problem solving and educational background

We have some experience of non-Type 1 problem solving in chemistry with Open University (OU) students. The student body of the OU is diverse, spanning extremes of ranges of age, education and background. An intriguing notion was to see if we could determine whether formal educational experience paralleled the ability to solve problems. Overall, there is evidence that OU students with recent experience of higher education study tend, at least to begin with, to perform to a higher standard overall than those without that recent experience[8] but would this be reflected in problem solving?

We divided a cohort (totalling 305 students) into three categories based on the England, Wales and Northern Ireland educational qualifications (Table 7). The Group 1 division separated the students

with no formal educational qualifications from those with any GCSE[*] or higher. The *same* cohort was divided differently to produced Group 2 which had as its lower educational population those students at least one GCSE grade and the higher educational population had specifically GCSE chemistry or chemistry within science or more advanced qualifications. The final Group 3 had the divide such that only Advanced-level (or higher) was included in the top division.

Table 7 Grouping of Open University students according to educational background.

Group	Lower educational population	Higher educational population
1	None	GCSE and higher
2	None plus GCSE	GCSE (chem) and higher
3	None plus GCSE plus GCSE (chem)	Advanced level and higher

Table 8 The relative performance of Open University students of different educational background on Type 4 problems.

Group	Low		High	
	% Score	Standard deviation	% Score	Standard deviation
1	63	14.0	65	13.2
2	63	12.6	62	13.7
3	66	12.2	63	13.9

The variations of student performance in Table 8 are statistically not significant. The performance on Type 4 problems does not seem to be affected by prior educational qualifications. (However, the initial performance of students shows significant variation on Type 1 problems with a correlation between prior educational level *and* how recent was that experience. These findings are consistent with those of Macpherson[9] who investigated the link between problem solving ability and cognitive maturity.)

---

[*] As with any formal, closed-book examination, it is not appropriate to test skills that have not been developed prior to the examination. Any move to include problem solving (Type 2 or higher) in the examination should be preceded with a familiarisation of problems of this type.

[*] The General Certificate of Secondary Education is a UK national examination. Students are normally aged 16 years and take up to 10 subjects.

Conclusions

This overview of first-year, university, chemistry examination papers embodies a number of limitations such as:

- The somewhat crude analysis of problem types and the ignoring of question types that cannot be categorised with this system.
- The semi-subjective mapping of questions to learning outcomes.
- The learning outcomes themselves, not all of which were defined clearly and related to the mode of assessment.

Nevertheless, the survey has thrown up a number of common features of examinations

- There is a mismatch between outcomes claimed and outcomes tested.
- Some outcomes are tested several times in the same paper and some omitted. (This situation is made worse where the paper embodies a choice of questions.)
- In the worst cases, students are able to achieve a pass grade with less than twenty per cent of the learning outcomes fully achieved.
- Questions that are easy to set and easy to mark tend to predominate.
- Claims for the assessment of problem solving should be viewed with suspicion without a clear idea of what constitutes a problem.
- Experience with simple algorithmic exercises is not an indicator of success with problem solving.

All the above findings effectively arose from our attempts to check that assessment was firmly embodied in our learning outcomes. The findings indicate that we have in some cases a way to go before we achieve a tight mapping between what we teach, what the students learn, and what we assess. A step forward would be to ensure that examination papers are subject to a simple learning outcomes analysis, which is then seen to be part of the total assessment of each module and programme. Learning outcomes, once defined, are capable of informing what we teach, what students learn and how it is assessed.

References

1 P. Race, *Designing assessment to improve physical sciences learning*, LTSN Physical Sciences Practice Guide, LTSN Hull, ISBN 1-903815-00-2, 5-6.
2 P. Ram, Problem-based learning in undergraduate education, *J. Chem. Ed.*, 1999, **76**, 1122.
3 Quality Assurance Agency UK, Southgate House, Gloucester GL1 1UB, www.qaa.ac.uk.
4 Quality Assurance Agency UK, Chemistry benchmark statement, www.qaa.ac.uk/crntwork/benchmark/chemistry.pdf
5 S. W. Bennett, *What do examinations really test?,* 15th International Conference on Chemical Education, Istanbul, 2004.
6 A. H. Johnstone, In C. Wood and R. Sleet (eds), *Creative problem solving in chemistry*, London, The Royal Society of Chemistry, 1993.
7 N. Reid and M. Yang, The solving of problems in chemistry: the more open-ended problems, *Research in Science and Technological Education*, 2002, **20**, 83.
8 *Interim results of student feedback from the 2003 National Student Survey*, Institute of Educational Technology, http://iet.open.ac.uk/nss/download/interim2.doc
9 K. Macpherson, Problem-solving ability and cognitive maturity in undergraduate students, *Assessment and evaluation in higher education*, 2002, **27**, 7.