

**A REFLECTIVE STUDY ON THE USE
OF CAA TO TEST KNOWLEDGE
AND UNDERSTANDING OF
MINERALOGY**

Alan Boyle

A Reflective Study on the use of CAA to Test Knowledge and Understanding of Mineralogy.

Alan Boyle
University of Liverpool
Department of Earth Sciences
University of Liverpool
4 Brownlow Street
Liverpool
L69 3GP

apboyle@liverpool.ac.uk

Abstract

Use of multiple choice question based computer aided assessment to assess level-one (first year) mineralogy produced a reliable assessment, though with rather poor scores. The use of negative marking contributed to this, and also drew negative comment from the student cohort. Reflection on these outcomes led to the use of multiple response questions, which performed better and did not encourage negative student feedback. CAA performance does not equate very well with practical coursework assessment. However, these two assessments are addressing different learning outcomes and so this disparity is not surprising. Statistical analysis suggests that these two forms of assessment give a truer indication of a student's ability when they are combined. It enforces the conclusion that appropriate assessment tools should be used for stated learning outcomes and that multimodal assessment is best.

Introduction

A common problem with assessment is that it comes at the end of a module, and has to be conducted and reported on in a short period of time before the examiner has to address other issues (e.g. the next semester or a period of research activity). The result of this is that all too often the outcomes of an assessment are not reflected on and as a result the assessment process is not improved. An additional problem is that the average university academic is unaware of research evidence on the validity of different assessment tools. This article describes assessment outcomes recorded over a three-year period within which changes were made in an attempt to improve the process. The article will first give some background to the module investigated, and then discuss assessment performance and the effects of changing some assessment tools.

Module Background

I took over the level-one module "ESGY105 – Minerals & the Microscope", comprising 12 lectures and 6 three-hour practicals, in session 1999-2000 following the death of the previous incumbent. Level-one mineralogy does not change greatly, so the material covered stayed the same, but the teaching, learning and assessment strategy did not. Material in lectures is now

delivered through PowerPoint making full use of multimedia for showing simulations, movies etc. and backed up by web resources. Assessment strategy was also re-assessed. To help understand the assessment strategy, it is useful first of all to summarise the intended learning outcomes for module: Students should:

1. know the properties of common rock-forming minerals;
2. understand common classification schemes for minerals and rocks;
3. understand how minerals may be interpreted to infer geological conditions and processes;
4. know how to use a hand lens and a petrological microscope;
5. acquire the skills needed to be able to recognise and make proper drawings of minerals in hand specimen and thin section.

The more practical aspects (4 & 5) are assessed via practical coursework and a short examination in the last practical [40% of the module marks] using a set of subjective grade descriptors. Prior to session 1999-2000, the more theoretical parts (1-3) were examined using a paper-based, 60 multiple-choice question (MCQ), four-items per question, examination [60% of the module marks], an objective method.

Session 1999-2000 Assessment

In session 1999-2000 the existing paper-based MCQ paper was converted, with little modification, to run as a CAA in January 2000 using the TRIADS engine (Boyle et al. 2000, Mackenzie et al. 2002). Some extra non-MCQ questions were added, taking advantage of facilities in the TRIADS engine, to test student understanding of classification schemes and their abilities to interpret data.

A typical MCQ from the year 2000 assessment is shown in Fig. 1. The MCQ

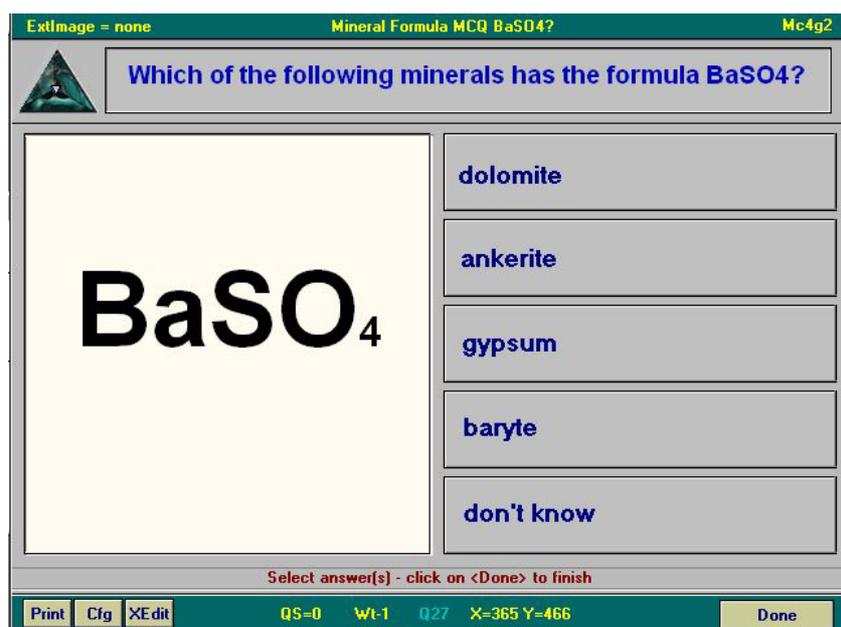


Fig. 1. Screen shot of typical MCQ in the year 2000 CAA.

asks about a mineral property (in this case, chemical composition) and gives four choices (the first 30 questions all addressed this aspect of mineral chemical composition in this style). Negative marking was employed to take account of guessing (so, the choice baryte scored 100%, the others scored -33.3%, “don’t know” scored 0%).

Item analysis indicated that the CAA performed well. Fig. 2 summarises the facility index (proportion getting the question right) and the point biserial index (PBI, did the right students get the question right) for the first 30 questions. A PBI for the keyed response of 0.40 and above are considered (California State Personnel Board 2002) to be very good (almost half the questions in Fig 2), while values in the range of 0.30 to 0.39 are good (a third of the questions in Fig. 2). A PBI of less than 0.30 for the keyed response suggests that the item is not doing an optimal job of discriminating between the better and poorer candidates (a fifth of the questions). Only two questions fall below a PBI of 0.2

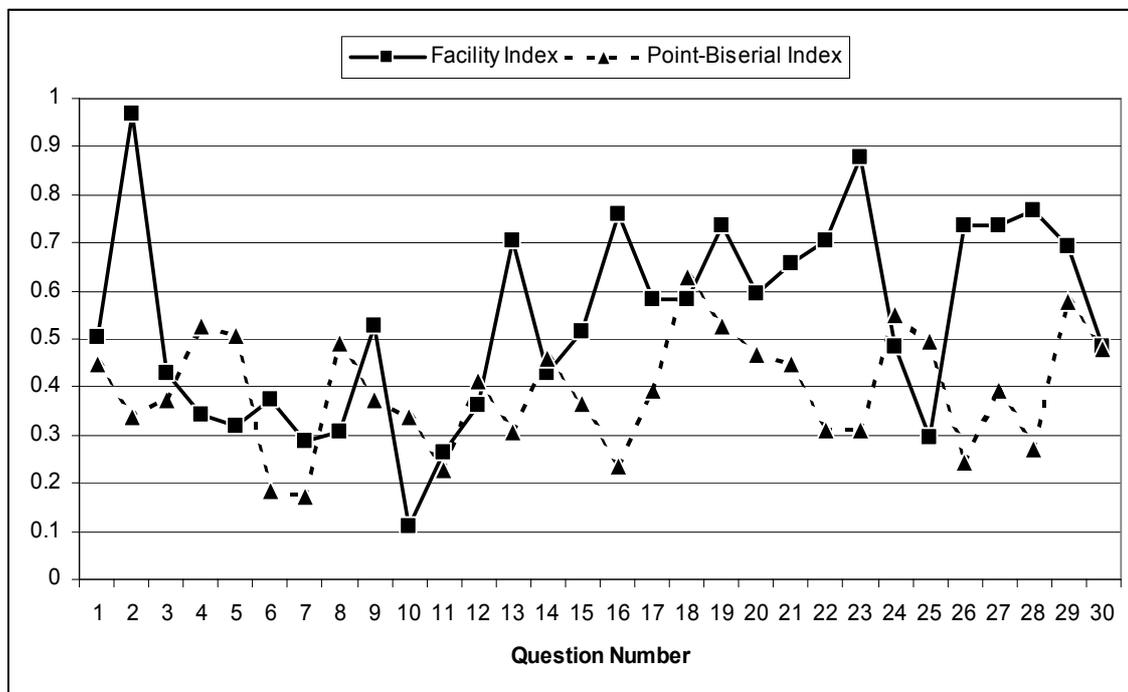


Fig. 2. Graph of Facility Index and Point Biserial Index for the first 30 MCQs in the 2000 test. Only the first 30 questions are presented as these questions form the basis of this paper.

	Practical	CAA
Avg	65.0	46.6
SD	10.2	16.9
Max	85.6	94.5
Min	6.9	12.4
Num	93	92

Table 1. Summary of practical and CAA marks for 2000. Num = number of students. One student missed the CAA.

However, the overall CAA scores were disappointingly low compared to the practical scores (Table 1).

Negative marking seemed to encourage significant non-answering of some questions, suggesting that it may have adversely affected the performance of some students. In addition, while student feedback completed after the CAA was very positive about the module as a whole, it complained bitterly about the unfairness of negative marking in the CAA.

Session 2000-2001 2001-2002 Assessments

In session 2000-2001, the issue of the potential effect of negative marking (and the student dislike of it) was addressed by altering many of the MCQs to multiple-response questions (MRQ). The MCQ in Fig. 1 is rephrased to be: "Which elements are normally present in baryte?", and the student then chooses from a matrix of 20 chemical elements commonly found in minerals (see Fig. 3). The question does not indicate how many correct buttons there are. Correct choices (Ba, S, and O in this case) score marks that will add up to 100% (33.3% for each) if all are correctly chosen, whereas incorrect choices score minus half the score of a correct choice (-16.7% in this case). A student selecting just the 3 correct items would get 100%, three correct and one incorrect 50%, just two correct 67%, two correct and two wrong, 33% and so on. TRIADS rounds the final score to an integer, and any overall negative score is zeroed. Each MRQ provides a range of marks, not the usual binary

42% done in 1% of time Q.36 Mineral Matrix

▲

Which of these elements are normally present in baryte?

Click on 'Done' when enough chosen.

Si	Al	Ti	Fe	Mn
Mg	Ca	Na	K	O
Pb	Zn	Cu	Cr	P
Ba	Cl	S	C	H

You may select more than one answer. Click on 'Done' when enough chosen

Quit Done

Fig. 3. Matrix style multiple response question (MRQ) version of the MCQ in Fig. 1.

range in MCQs. For some minerals there are elements that may be present, but are not “normally” so. In these cases, choosing that element would score zero (i.e. be neutral)

Analysis of the test outcomes indicates that the overall scores were more in line with expectation using MRQ than MCQ with attendant negative marking, though the CAA marks are still lower than the practical marks (Table 2). Students did not complain about this question style.

Comparison of MCQ and MRQ performance

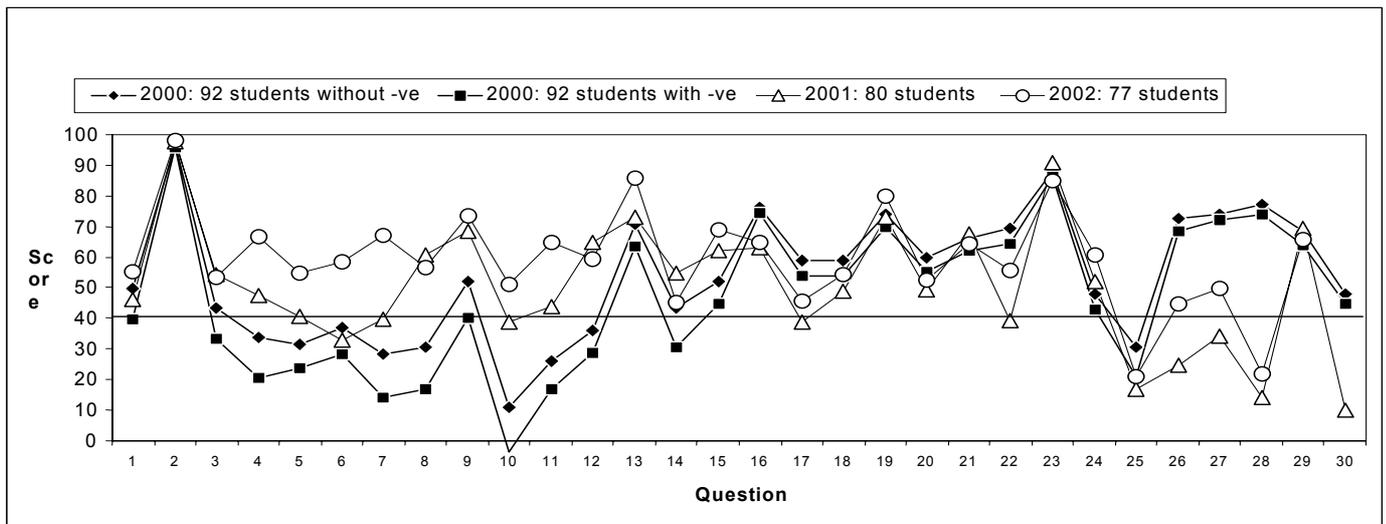


Fig. 4. Comparison of 30 mineral chemical composition question scores in 2000 (MCQ with and without negative marking) and 2001 and 2002 (MRQ).

Q1 = albite; Q2 = quartz; Q3 = anorthite; Q4 = orthoclase; Q5 = muscovite; Q6 = diopside;
 Q7 = hypersthene; Q8 = biotite; Q9 = olivine; Q10 = chlorite; Q11 = garnet; Q12 = hornblende;
 Q13 = haemetite; Q14 = ilmenite; Q15 = magnetite; Q16 = chromite; Q17 = rutile;
 Q18 = chalcopryrite; Q19 = galena; Q20 = sphalerite; Q21 = pyrite; Q22 = aragonite; Q23 = calcite;
 Q24 = dolomite; Q25 = siderite; Q26 = magnesite; Q27 = barite; Q28 = fluorite; Q29 = halite;
 Q30 = apatite.

Figs. 1 and 2 represent different ways of assessing the same information,

	Practical 2001	CAA 2001	Practical 2002	CAA 2002
Average	68	55	67	52
SD	10	16	9	16
Max	89	94	91	88
Min	39	24	46	21
Num	82	80	79	77

Table 2. Summary of scores for practical and CAA components in Years 2001 and 2002. Num is the number of students. For both years two students missed the CAA.

namely what is the chemical composition of baryte. Fig. 4 summarises the average score for the 30 mineral composition questions run in 2000, 2001 and

2002. There are two sets of data presented for 2000, the raw data including the negative marking (i.e. an incorrect answer gave a question score of -33% thus lowering the average question score) and modified data in which all -33% scores were zeroed. The modified data obviously give higher average question scores than the raw data. 40 % is the institutional pass mark.

It is clear that many of the MCQ versions of the questions gave fail marks (even for the modified set), whereas the MRQ equivalents generally did not. The questions are arranged in mineral groups. Thus questions 1-12 are all for silicate minerals, and the MRQ versions generally gave higher scores. Questions 13-17 are for oxide minerals, 18-21 for sulphide minerals and 22-26 for carbonate minerals. All of these questions perform similarly for MCQ and MRQ, with the exception of question 26. This question performs like questions 27, 28 and 30 which produced much higher scores as MCQs than as MRQs. Why should changing the format of a question change the score from a first class mark (average over 70%) to a low or even fail mark (below 40%) in this last group of questions? Table 3 summarises the four questions, their items and the correct answer. For questions 26-28 the answer is largely indicated by the presence of elements Mg, Ba and F respectively. The question gives a big clue as to the correct answer.

Comparison of CAA and practical scores

Question stem	Items	Answer
Q26 Which mineral has the formula $MgCO_3$?	siderite magnesite calcite dolomite	magnesite
Q27 Which mineral has the formula $BaSO_4$?	dolomite ankerite gypsum baryte	baryte
Q28 Which mineral has the formula CaF_2 ?	halite pyrite fluorite aragonite	fluorite
Q30 Which mineral has the formula $Ca_5(PO_4)_3(F,Cl,OH)$	halite fluorite aragonite apatite	apatite

Table 3. Summary of question stems, items and correct answer for 4 MCQs that were well answered in 2000, but poorly answered as MRQs in 2001 and 2002.

Fig. 5 summarises the CAA and practical scores for all students taking module ESGY105 over the three years considered. There is a poor positive correlation. The objective CAA scores show a characteristic wide spread of values whereas the subjective coursework scores show a characteristic narrower spread of values. Year 2001 has a marked “hump” in the distribution

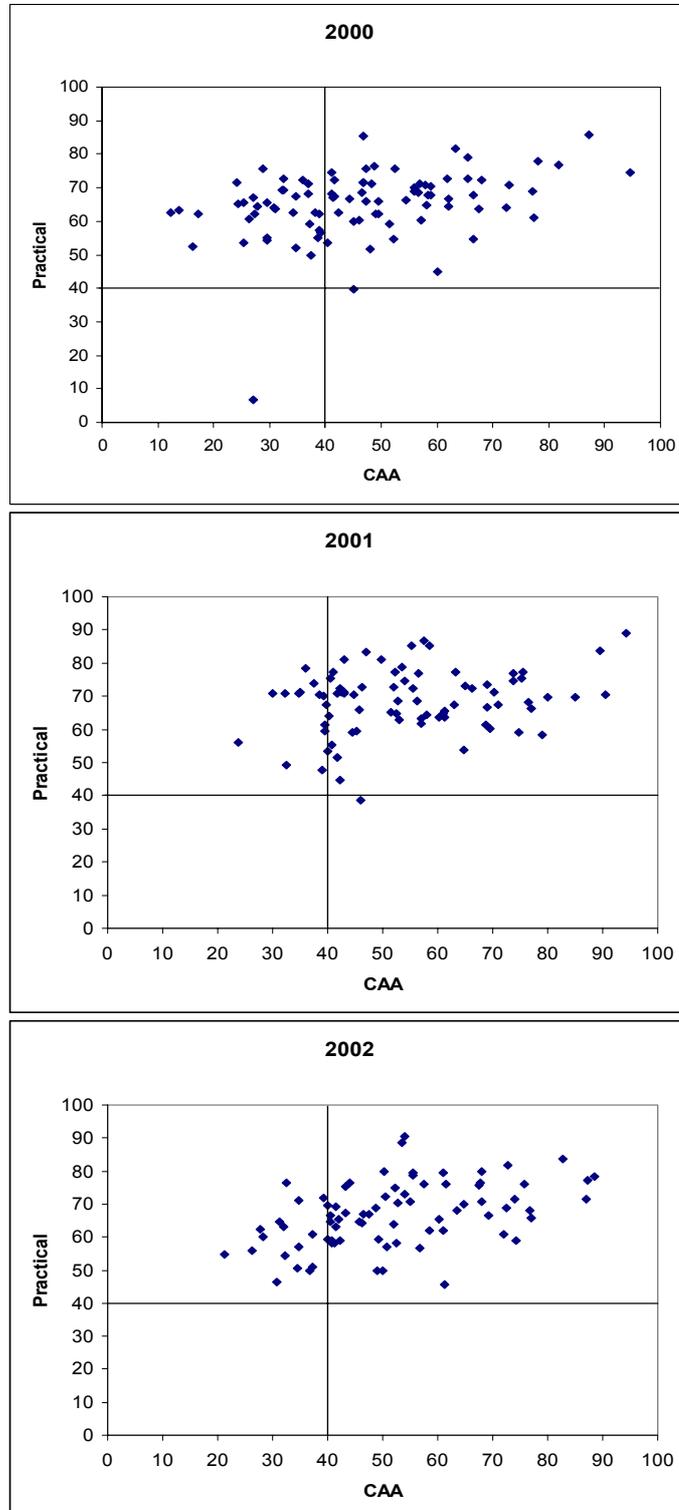


Fig. 5. Graphs of CAA and practical marks over a three-year period. 40% represents a pass mark.

of data points relating to students who scored relatively highly in the practical components but less well in the CAA.

The same feature is present, though less well developed, in the other two years. Note the higher density of data points towards the low-CAA mark end defining a steeper trend than the overall trend of data points for examination years 2000 and 2001.

What is the significance of this? The two assessments are not only assessed in different ways, they are assessing different things. The practical assessments are testing the student's ability to observe record and interpret real objects seen with a hand lens or with a microscope. They very largely test skills-based outcomes and the scores are based on a set of subjective grade descriptors. The CAA is more concerned with testing knowledge and to some extent understanding and the scores are objective. Some students are clearly very skilful at observing and recording, but less good at remembering mineral properties and how minerals are classified. Similarly, there are students who have a good knowledge and understanding of mineralogy, but are not very good at observing and recording. Given the intended learning outcomes of the module, both of these assessments are necessary

	2000	2001	2002
Practical	0.450	0.512	0.643
CAA	0.512	0.568	0.749
Module	0.576	0.652	0.808

Table 4. Pearson correlation coefficients for component practical and CAA scores as well as overall ESGY105 module score against performance in other modules taken the same session. For 2000 and 2001 this is 7.5 modules, whereas for 2002 it is only 3.5 modules because second semester marks are not available at the time of writing. All correlations are at the 99% confidence level.

Comparison of CAA, Practical and Module Scores with Student Overall Performance

ESGY105 is a half module and, as such, represent 1/16 of a student's year of study at Liverpool University. Table 4 summarises Pearson correlation coefficients for CAA, practical and overall ESGY105 module scores against overall student performance in other modules. For 2000 and 2001 this is against the average mark for 7.5 modules, for 2002 it is against the average mark for 3.5 modules since the marks for the second semester have not yet been finalised. It is notable that:

1. all correlations improve with time;
2. Module correlation is always better than CAA, which is always better than practical.

The first of these may of course represent improved teaching over time, particularly for the practical component. However, I suspect that the improved correlation of the CAA component reflects the changes made in the

assessment to replace MCQs with MRQs and remove carry-over negative marking, together with other reflective modifications of questions to remove perceived ambiguities. The fact that the overall module score has the best correlation is interesting. Pearson Correlation coefficients for Practical:CAA over the three years are 0.348, 0.270 and 0.489 at the 99% confidence level. The fact that the overall module score always has the best correlation with performance in other modules suggests that the multi-modal assessment strategy used is actually best (as advocated by Fehring 2001 and Sherman 1997), despite the poor correlation between the two components themselves. Assessments based solely on CAA should perhaps be viewed with caution, and care should be taken when relating the scores they give to scores from necessarily more subjective assessment methods.

Conclusions

Multiple choice style questions are easy to implement and can produce reliable assessments. However, if carry over negative marking is used to overcome the guess factor, the resulting scores are rather low and students dislike them leading to negative feedback in module review. Reflection on this problem indicates that multiple response questions with internal negative marking provide an attractive alternative to standard multiple choice questions, and perhaps provide a better assessment tool. Finally, the appropriate assessment tool must be used. For this module, CAA alone would not be a good assessment tool because it would not assess practical skills. In addition, while CAA scores correlate better than practical scores with individual student performance in other modules, suggesting they may be better measures of student ability, when combined into a module mark they correlate still better. The implication, in this case, is that multimodal assessment provides a better indication of overall student performance.

Acknowledgements

Jim Marshall is thanked for reading through an early draft and providing very useful comments. The ciad team at the University of Derby are thanked for continuing to provide and support an excellent engine (TRIADS) for delivering CAA.

References

Boyle, A. P., Edwards, D. J., Mackenzie, D. M., Mills, B., O'Hare, D., Morris, E. C., Paul, C. R. C., Wilkins, H. & Williams, D. W., (2000). Developments in on-line assessment - experiences and evaluation of using TRIADS and its potential for assessment in Electrical Engineering. *International Journal of Electrical Engineering Education*, **37**(1), 74-85.

Bull, J. and C. McKenna (2001) *Blueprint for Computer-assisted Assessment*, CAA Centre, University of Luton.

California State Personnel Board (2002) *Interpreting Item Analysis Data*
<http://www.spb.ca.gov/Interpret-item-analysis.pdf> (21 May 2002)

Fehring, H. (2001) *Literacy assessment and reporting: changing practices*
<http://sece.eu.rmit.edu.au/staff/fehring/irish.htm> (21 May 2002)

Mackenzie, D. M., O'Hare, D., Paul, C. R. C., Boyle, A. P., Edwards, D. J., Williams, D. & Wilkins, H., (2002). Assessment for Learning: the TRIADS Assessment of Learning Outcomes Project and the development of a pedagogically-friendly computer-based assessment system. In: *Developments in online assessment* (ed O'Hare, D.), SEDA.

Sherman, L. (1997) *Research review*
http://www.nwrel.org/nwedu/fall_97/article6.html NW Education Magazine Fall Issue (21 May 2002)