

BIOSCOPE THE ASSESSMENT OF PROCESS AND OUTCOMES USING THE TRIADSYSTEM

Sarah Maughan and Don Mackenzie

BioScope

The Assessment of Process and Outcomes using the TRIADSystem

Sarah Maughan
University of Cambridge Local Examinations Syndicate
1 Hills Road
Cambridge CB1 2EU
maughan.s@ucles.org.uk

Don Mackenzie
Centre for Interactive Assessment Development
University of Derby
Kedleston Rd.
Derby DE22 1GB
d.mackenzie@derby.ac.uk

Abstract

An interactive assessment using a simulated biological microscope to test process as well as outcomes is described. The initial results of teacher and student feedback from trials in international schools are reported and evaluated.

Introduction

In 2002 the University of Cambridge International Examinations (CIE) commissioned the Centre for Interactive Assessment Development (CIAD) at the University of Derby to create a biological microscope simulator (BioScope) for use on their teacher support web site. This simulator was an extensively re-coded version of a similar petrological microscope simulator that was developed in 1995 for the DIADS pre-cursor to the TRIADSystem. The text content, functionality and slide sets used in the BioScope are defined via external XML control files with JPEG graphics representing the slide images in the various modes. Controls include focus, magnification, movement, eyepiece graticule, stage scale and a variety of pre-defined filters. After successful trials with teachers, the BioScope has been recently published by Cambridge University Press, populated with CIE learning resources and slides suitable for IGCSE/ O Level and A Level.

Building on this development, CIE commissioned CIAD to develop a short practical assessment in Histology where the candidates would use the BioScope to investigate a number of plant sections, make measurements and draw conclusions concerning their ecological environment. The assessment was designed to test the same assessment objectives as are currently targeted in the microscopy component of an AS Level Practical exam. However, the aim was to enhance the assessment by attempting to measure the process that the students used to reach their answer, rather than just the outcome itself.

The BioScope assessment has recently been trialled in a number of international schools and the first part of the evaluation of these trials, including the qualitative feedback from students, is presented together with a demonstration of its functionality.

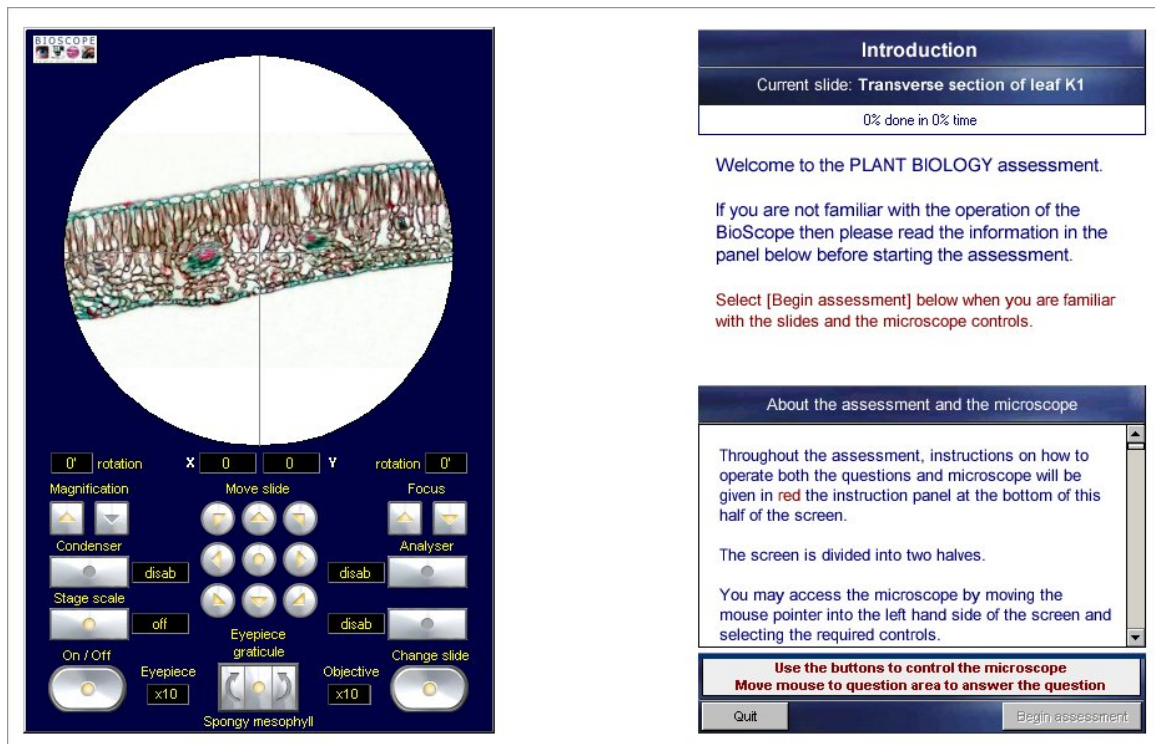
BioScope in TRIADS

The semi-open coded nature of the TRIADSystem makes it especially amenable to the construction of assessments that involve scenarios, simulations, bespoke scoring, tutorials and/or extensive feedback options. The BioScope was therefore embedded in the TRIADS engine to produce an assessment that allows the candidate to use it as an investigative tool to answer the questions in the same way as they would use a conventional microscope.

There are many advantages to using a simulated microscope for practical assessments in science, not the least of which is that all candidates are investigating the same slides where the quality can be guaranteed and with identical microscope functionality. The expense of producing and quality assuring multiple sets of slides is removed and the marking can be more precise. In traditional practical assessments involving the use of a microscope, the examiner really has no idea of how proficient the candidate is at using the instrument. In the BioScope, all areas of each microscope slide are invisibly 'hot-spotted' and the BioScope assessment produces a continuous record of the candidate's use of the microscope. In some questions the appropriate use of the microscope forms an integral component of the scoring and where questions are focused on a particular feature on the slide, the system checks that the candidate is actually examining or has examined that feature when submitting their answer.

Before starting the assessment, the candidate is allowed some time to familiarise themselves with the BioScope. (Figure 1)

Figure 1 BioScope familiarisation screen

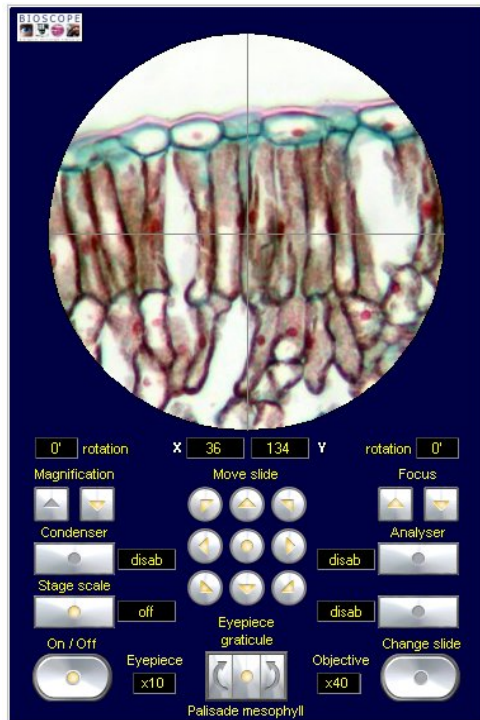


Examples of questions

In traditional practical examinations where candidates are required to draw and comment on or label particular areas of the sample, the examiner has no way of checking that the drawing represents the actual sample under the microscope or whether it is an example memorised from a text book by the candidate.

In the BioScope assessment, the candidate is required to select a [Draw] button when they have chosen the area to be drawn and the system records a snapshot of the slide and microscope settings. (Figure 2) The individual candidate record file is output in HTML format so that the snapshot can be embedded into the printout for comparison with the drawing by examiners during the marking process. The system allocates scores for correct focus, magnification and slide position to contribute to the marks allocated for the candidate's drawing and labeling.

Figure 2 Draw the sample question



1. Draw the palisade mesophyll

Current slide: Transverse section of leaf K1

0% done in 1% of time

The slide shows a stained, transverse section of a leaf.

Make a labelled, high-power drawing on the paper supplied to show the detailed structure of three adjacent cells from the palisade mesophyll.

Click the [Draw] button when you have selected the area and are ready to start drawing.

DRAW

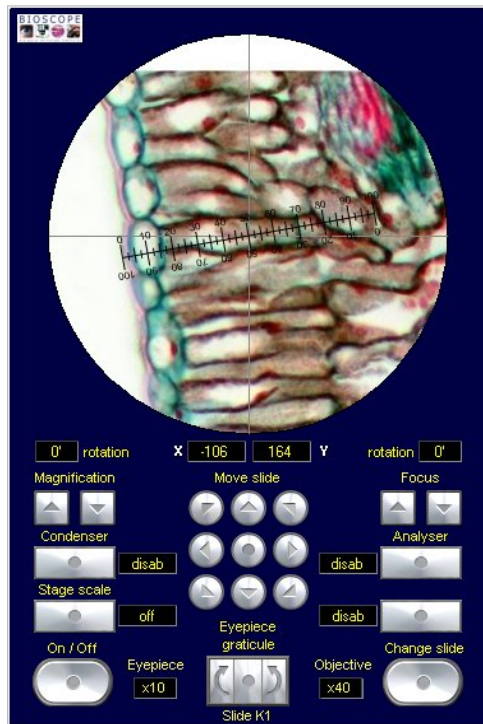
Use the buttons to control the microscope
Move mouse to question area to answer the question

Quit

Candidates are commonly required to make quantitative measurements of specimens during practical examinations and in the BioScope assessment they are required to measure the average thickness of Upper Epidermis cells using an eyepiece graticule that can be rotated, together with a stage scale to calibrate the graticule in absolute units and an on-screen calculator. All operations are manipulated by means of buttons on the microscope controls so that the candidate can measure a number of cells using the graticule in graticule units (Figure 3), then convert these measurements into absolute units (mm) by calibrating the graticule against the stage scale for the magnification used.

This is an operation that candidates traditionally find quite difficult, even at first year undergraduate level. During these operations, the BioScope engine checks that the candidate has the graticule in the correct orientation to make the measurements, that the minimum number of measurements has been made to achieve a realistic average, that the individual measurements are within the acceptable range and that the average has been correctly calculated with respect to the candidate's measurements. Each stage in the conversion of the average thickness in graticule units to the absolute average thickness and choice of units is also checked by the system and scored appropriately. In this way, marks can be awarded for the proportion of the process that is correct even if an error along the way results in an incorrect final answer.

Figure 3 Making measurements



3. Measure u.epidermis in gu

Current slide: Slide K1 subsection

29% done in 9% of time

Use the eyepiece graticule to measure the average THICKNESS of the UPPER EPIDERMIS including THE CUTICLE, in graticule units.

Make AT LEAST THREE measurements and record them in the spaces provided below.

Calculator

16.3333333

1	15	5
2	18	6
3	16	7
4		8

Average thickness is: 16.34

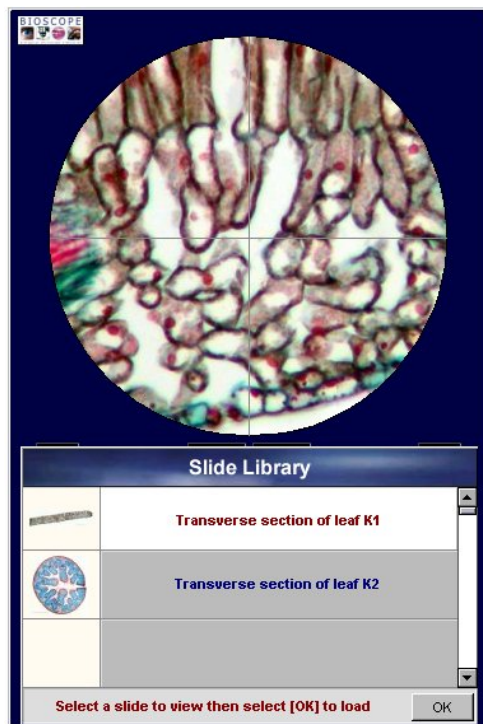
Graticule Units

Click on entry box, type answer, press 'Enter'
Select [Submit] to finish.

Quit Clear All Clear Submit

Later in the assessment the candidate is required to use the BioScope to compare the features of two samples with respect to their habitat. The BioScope facilitates this by the provision of a Slide Library from which the candidate can select the slides to be displayed and investigated. (Figure 4) Where it is required that the candidate examines a particular feature of a sample in order to come to a conclusion, there are inbuilt checks to ensure that the candidate has actually investigated that feature. If the candidate has not actually centred the microscope on the required feature then they are prompted to do so when confirming their answer and marks are only awarded if they can actually locate the required feature.

Figure 4 Use of the slide library for sample comparison



7. Comparison of Stomata

Current slide: **Transverse section of leaf K1**

86% done in 19% of time

Compare the stomata of slides K1 and K2 using the [Change slide] button to toggle between slides on the microscope.

How would you best describe the contrast between the two specimens?

Position of stomata	only in		in both
	K1	K2	
all over surface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
exposed directly to environment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
on one surface only	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
not exposed directly to environment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Protection from environment	K1		K2		both
	protected in deep pits or folds	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
protected by hairs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
protected by thick cuticle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
protected by thorns	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
no protection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Airsapces	K1		K2		both
	stomata open onto a network of airsapces	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
stomata open onto small enclosed airspace	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
stomata do not open onto airspace	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Use the buttons to control the microscope
Move mouse to question area to answer the question

Quit

Throughout the test there are examples where the candidate's answer to a previous question is carried forward to subsequent questions. In cases where this will impact on the candidate's ability to answer aspects of the subsequent question then corrected information may be provided but the candidate cannot return to edit their previous answers.

Evaluation

The BioScope assessment has recently been trialled in a number of international schools and the first part of the evaluation of these trials, including the qualitative feedback from teachers and students, is presented. For the pilot, schools were selected that have students entered for the AS Biology examination in June 2004. This exam has a compulsory practical component. Approximately 15 schools were targeted with the hope of completing the pilot with over ten schools, in the end, results were received from 11 different schools in countries including Malaysia, India, Pakistan, Kenya and Spain.

The aims of the pilot were to:

- 1 investigate whether the test as a whole, and the individual items within it, meet the requirements of a traditional test analysis,
- 2 collect views of teachers and students on the tests,
- 3 compare student performance on the tests with their performance on existing examinations.

The first two aims will be evaluated in this paper, but the data for the third area will not be available until later in 2004 (after the June 2004 examination session has taken place).

Once schools had agreed to take part they were sent the assessment on CD and sufficient questionnaires for each student and for the teacher. The tests were sat in the schools in the way that was most appropriate to them, in most cases this was all students in the class together, using the computer lab. After the assessment had been sat results files were returned to Cambridge. For the purpose of the analysis the results file with marks at the item level was used to undertake a traditional test analysis. The results of the item analysis are given below, as are the results of the questionnaire analysis.

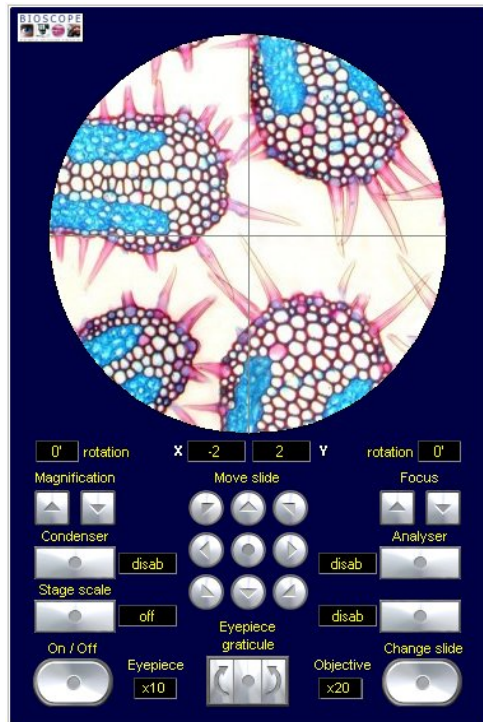
Item Analysis Results

Overall the test is measuring approximately in the correct difficulty range, with total marks ranging from 1 to 26 out of 27 and the mean mark of 63.7% is similar to that of the paper-based equivalent, for example the June 2002 equivalent component had a mean mark of 63%'. It will be interesting to see how these results compare with those from the live exams for the same group this year.

The results of the analysis by gender show very little difference in performance between boys and girls with the mean marks for boys and girls being 65.1% and 63.1% respectively.

In terms of individual items the test appears to be fairly well ordered with item facilities decreasing from 0.91 to 0.32 across the test. Only question 5 seems to be in the incorrect position with a facility of 0.83, but it also has a very low correlation with test total (0.25), which could be partly to do with the high facility value. The question is shown below (Figure 5) and it is a single mark question (compared to the other questions which are 6, 4, 6, 3, 4 and 3 marks). It could be that this question has a low correlation because it is assessing the students deduction of information based on their knowledge, compared to all other questions in the assessment which test practical skills.

Figure 5 Assessing straightforward recall of information in Q.5



5. Habitat of K2

Current slide: Transverse section of leaf K2

1 mark 57% done in 14% of time

Look at the transverse section of the leaf K2.
This plant is adapted to a particular habitat.

What type of habitat is this?

Press [Enter] to complete your answer.

Click on entry box, type answer, press 'Enter'
Select [Submit] to finish.

Quit Delete Submit

On the whole we did not find the large differences between boys and girls on individual items, as we did not find differences across the whole test. There were three questions on which boys did markedly better than girls: 2, 3 and 6, although there is no obvious reason why this should be the case.

Analysis of Teachers' Questionnaire

Completed questionnaires were received from ten teachers, all of whom rated their computer skills as 'neither good nor bad', 'good' or 'very good'. The teachers all felt that at least some of their students found the Bioscope easy to use, but 9 of them felt that it could be made easier in some way. Their comments focused particularly on more practice prior to the exam, so although the practice area was provided, most teachers did not seem to find this sufficient. This is an area that has been considered in the development phase of the Bioscope so versions are available for use during the learning, both a small sample of tasks on the CIE Teacher Support Site and a much wider variety on the version published by CUP. It is expected that if the Bioscope becomes used in 'high stakes' exams that the learning tools will be used beforehand. The second comment, mentioned by two teachers, was the ability to go back and review answers. Obviously in paper-based exams this facility is available to students, and they are taught throughout their schooling

to use this option. Both students and teachers found it difficult to have this option removed, and this is something we need to consider if we introduce simulations into high stakes examining, it is likely that we will include the option of going back to questions in future iterations of the Bioscope and this will have implications for the questions that can be asked or the way in which responses are stored and marks allocated. Interestingly the ability to page between questions was one of the main points made by university students in the evaluation of the 'Assessment of Learning Outcomes' project (Mackenzie, et.al. 2004).

The teachers found the instructions for using the Bioscope clear, the quality of the graphics clear, on the whole, and the timing of the assessment about right. These findings would suggest, along with the item analysis above, that the assessment has been pitched at the right level for the students, and that their performance has not been affected by extraneous factors.

When asked if teachers felt that the Bioscope had provided a fair assessment of the students' ability in Biology three teachers answered 'yes', five 'no' and four 'not sure'. These numbers are perhaps less positive than intended as the comments provided alongside them focused in two cases on the lack of a 'wet practical' in the assessment, which is the other half of a traditional AS Level practical exam. This omission was intentional at this stage and the wet practical is likely to be the area of phase 2 of the development. However, the majority of comments expressed a concern about the students not undertaking 'real' practical work. 'The skill work and experience of handling an actual leaf or flower and staining the section is missing.' 'It will never substitute for actual practical skills.' This is an area that needs to be investigated further with teachers, and communicated effectively. At CIE we believe that practical work is a key part of any Science course, however, for various reasons, it may not be the best form of assessment of the skills involved, with reasons including comparable access to facilities in schools, manageability of assessment and standardisation of slides etc. However, acceptance of new assessments by teachers will be important for future developments.

Only one teacher commented that computer skills may have an impact on results.

Analysis of Students' Questionnaire

Completed questionnaires were received from 164 students. 88% of the students who participated in the Bioscope pilot claimed their computer skills were 'neither good nor bad', 'good' or 'very good' and only 5% found the instructions about how to use the Bioscope either 'unclear' or 'very unclear'.

When using the assessment both the simulation and questions were open at the same time on two halves of the screen. Students were asked how easy they found it to move between the two halves and only 7% found it 'not easy' or 'very difficult'. These responses from the students again reinforce the view

from the teachers' questionnaire that students' performance should not have been affected by extraneous factors.

The majority of the problems mentioned by the students appeared to be caused by speed of the computers, especially at high slide magnifications, so it is clear that minimum computer specifications must be given for live examinations. However there was one comment about a closer integration of the two halves *'It would be better if the question was on the microscope slide itself.'* although not all students agreed with this view: *'The questions were clear and well away from the slide so there was no interference.'* From the pilot it appears that students definitely prefer the two parts visible at the same time, but closer integration than that is more controversial.

For the different tools in the simulation, the majority of students found these 'easy' or 'neither easy nor difficult' to use, with the percentages of students saying this for each tool being as follows:

Magnification	91%
Focus	93%
Move slide (NB there were 21 comments that this was slow)	90%
Stage scale	74%
Eye piece graticule	78%
Change slide	98%

It is interesting that the two tools that were hardest to use were the tools that the students have most difficulty with in real microscope work.

Two students commented that they found the graticule difficult because they did not understand it, this may be an example of where using the simulation provides an opportunity to use tools that are not widely used or to give additional practice for areas that students have difficulty with. One student commented that it is easier to use the graticule in the simulation than on a microscope, so it may be possible to use this for scaffolding the learning as part of the course.

The different item types were generally found easy to use, with the hot spot question (question 2) being found 'easy' or 'neither easy nor difficult' by 83%, the one word response again by 83%, the selecting characteristics from a list, 80%, radio buttons, by 72%. Where question types were not found easy it tended to be caused by difficulties with the task rather than the tool although some comments were about the function of the tools themselves:

Tool	Task-related comment	Tool-related comment
Eyepiece graticule	'did not understand how to do it'	'I didn't know how to place the graticule in the right direction'
Stage scale	'not taught how to use the stage scale' 'the tools were easy to use, but I found the question difficult'	'thick lines make it difficult to calibrate'
Radio buttons	'no stomata could be seen in K2, only one seen on K1'	'checkboxes are easier since you can eliminate options'

One of the key advantages of the Bioscope is that it can mark the work as the student completes each task and can provide immediate feedback to the student, in this case feedback was only provided on exiting the test. 68% of students commented that this feedback was helpful, although 18% found this 'neither helpful nor unhelpful' and 11% found it 'unhelpful'. For high stakes examining, at the end of which the students cannot do anything about their performance, the benefit of feedback may not be so great. This may be an area for development of assessment for learning materials using the Bioscope, which the student can use as part of a course to improve their skills.

The students were asked whether they felt this use of simulation would give a fair assessment of their Biology skills and knowledge and 36% said it would, a further 29% were unsure and 35% said no. Examples of the comments given by the students are below:

Comments against the use of the simulation:

'A practical exam should be done in a lab and the experiment should be carried out.'

'Doing an exam on computer is not actually a test of practical skills.'

'Some students are better at using computers than others.'

Comments in favour of the use of the simulation:

'Better because complete control is given to the student and all students have the same clear slide.'

'It could allow for more of the syllabus to be covered since manual practicals take a lot of time.'

Like the teachers, two students commented on not being able to go back through the questions to check the performance. For future pilots and for live

exams it should be recommended that questions are set up so you can go back to them. This is what we generally encourage as good test practice.

As to whether they would choose to take their exam in this way, 39% prefer the simulation, 20% prefer a paper-based alternative and 28% would prefer a practical exam. This means, that of the three methods, the simulation was the most popular, although it was very varied. Comments included:

'To develop skills required later in life. Practicals are important as many biology students are future doctors.'

'Find it easier to apply my knowledge in paper-based exams.'

'Accurate results and the same slides and apparatus for all students.'

'It makes the assessment interesting and less tension is created than in the practical examination.'

There were many suggestions as to how the simulation could be improved and these have been categorised below:

Instructions (either more or fewer or clearer): 10

Clearer slides or higher magnification: 15 (plus one request to see the slide on the whole screen)

Return to previous questions: 3

Speed and ease of navigation: 3

A number of students made comments about computers freezing and students losing their work, this is linked to a minimum specification for computers used, and also to a requirement of some form of caching of results during the exam. This will need to be addressed in future applications of the software.

Conclusions from the Bioscope Pilot

The results of the pilot were generally very positive with many students particularly being ready to accept this form of assessment as an alternative to the current examinations, choosing it over both paper-based exams and practicals. It was clear that there were some issues with the software and the way it was set up; there were issues with being able to go back and review work, speed of functioning and loss of work when a computer froze, and there was the issue of familiarity with the software. It is clear that these issues will need to be addressed prior to the Bioscope being used more widely in high stakes exams.

Future Development

Further analysis will be undertaken of the performance of the assessment, once results are available for students in their final exams in June. Once the full evaluation results are available decisions will be made regarding the future development of this and other simulated environments for use in high stakes examinations. Based on current findings it is likely that a phase 2 development will be undertaken focusing on the wet practical that makes up the other half of an AS Level practical exam, and a further development of microscopy tasks could be undertaken using new slides and implementing some of the lessons learnt from this pilot. In the UK the Qualifications and Curriculum Authority has set ambitious targets for the introduction of computer-based tests at GCSE level, if these are met assessments such as this use of the Bioscope may be seen as part of high stakes exams over the next few years.

The BioScope is currently being extended to provide the functionality of a petrological microscope, to include simulated slide rotation, polarizing filters, sensitive tints and Becke Lens, for geological applications. This version has already been used to deliver an assessment in Forensic Mineralogy at the University of Derby. Thus, its development has come full circle from its original design but the system is now in a more sophisticated format and integrated with full user-tracking to facilitate the assessment of process as well as outcomes in a wide variety of applications.

Reference

Mackenzie, D.M., O'Hare,D., Paul,C., Boyle,A., Edwards, D., Williams,D. & Wilkins, H. (2004) Assessment for Learning: the TRIADS Assessment of Learning Outcomes Project and the development of a pedagogically friendly computer based assessment system, p.22. In O'Hare,D & Mackenzie, D.M. (Eds) *Advances in Computer Aided Assessment*, SEDA Paper 116 pp11-24. Staff and Educational Development Association Ltd., Birmingham.