Maths-CAA Series
June 2003

Incorporating Assessment into an Interactive Learning Environment for Mathematics

Contents

Incorporating Assessment into an Interactive Learning Environment for Mathematics

Manolis Mavrikis and Antony Maciocia
School of Mathematics, The University of Edinburgh, Scotland
Email: M.Mavrikis@ed.ac.uk, A.Maciocia@ed.ac.uk

Abstract: In this article, after briefly describing a web-based Interactive Learning Environment (WALLIS) and the reasons for incorporating assessment into it, the approach to implementing formative CAA is described. This has recently been pilot-tested with a first year honours group of students. The pilot test results indicate that the introduction of the new method of delivery had no adverse effect on students' performance. In fact, their willingness to use the system, their results, and particularly their comments from interviews that were conducted were satisfactory enough to want to continue to develop further materials and continue to integrate CAA into the system. In order to do this effectively, recommendations that occurred from our pilot-test and from relevant literature in the field are described. Finally, in order to be able to share the experience and content the issue of re-usability and some possible directions that could be taken are considered.

1. Introduction

The growing deficiency in mathematical skills amongst science and engineering students is a recurring theme within LTSN articles and related bodies (for example see [1, 2]). The increased intake and the diversity among students make additional support (tutorials, formative assessments, accurate feedback) difficult not only for non-specialist but also for honours courses. These problems have significant negative pedagogical effects and lately universities have taken action to improve the situation by looking for e-learning solutions. Similarly, the above School under a KETF-SHEFC grant is investigating student support through a web-based Interactive Learning Environment (ILE), named WALLIS (after the famous 17th century mathematician) [3]. This system is briefly described in the following sections.

Recognising that, in order to be effective, applied technology must be embedded in the whole range of the student learning experience, it was decided to extend WALLIS to include more formally assessed parts. Assessment, apart from organisational issues (such as regulating teaching and informing the institution on students' progress), is an integral part of the students' learning experience. Previous articles in the mathematics-caa series [for example 4,5] or other related publications [such as 6] have described in detail not only its practical implications, such as the reduced load to mark assessment, but also the pedagogical ones such as providing students with accurate and timely feedback. There were further reasons to test the on-line delivery of assessment. Firstly, there is the need to explore both the benefits of using such an approach as well as the students' perceptions of Computer Assisted Assessment (CAA) in our context. Additionally, relevant research [for example 7] has noted that "deep learning" involves more than motivating materials or new ways of delivering assessment but rather the wider student learning experience. Employing the same means for both delivery and assessment brings assessment closer to the context in which students are studying and (as was proven by the pilot-test) motivates students to use and learn from the on-line material.

On the other hand, there was the desire to integrate assessment that would potentially have the ability to test higher mathematical skills, provide accurate feedback, target students' misconceptions, help them learn more efficiently, while being easily authored, maintained and based on sufficiently general technology to be able to use it in other topics than the one to be pilot-tested; hence the approach described in this article.

2. The Design of WALLIS and its Application

2.1 The main framework

After exploring many possible solutions, and various pilot tests it was decided to avoid limitations that off-the-shelf software packages have (particularly related to mathematical input and notation) by employing our own Java Server (Tomcat) with the hope that such a system will be more 'future proof' but also having in mind the long-term goal of the project to ease the authoring of material by lecturers themselves. The system is delivered through a web browser and consists of several pages containing theory, examples, self-practice activities and questions. Following a more constructivist view and to actively engage learners with the materials, the view was taken that simply reading through some on-line text resulted in passive learning. So it was decided that interactive activities would be needed to overcome this problem. While students work through the various pages they can interact with the insets to see various items such as a different example, some special cases of a formula, animations on how to do things. This is material that cannot, in general, be seen in a static book or web page. On the other hand, even interactive activities can often portray a different meaning to a student than the designer intended. Recognising the fact that they must remain involved, active, and challenged to think about and learn the presented material [8], a feedback frame is employed (see figure 1) which delivers context sensitive help relative to students' answers.

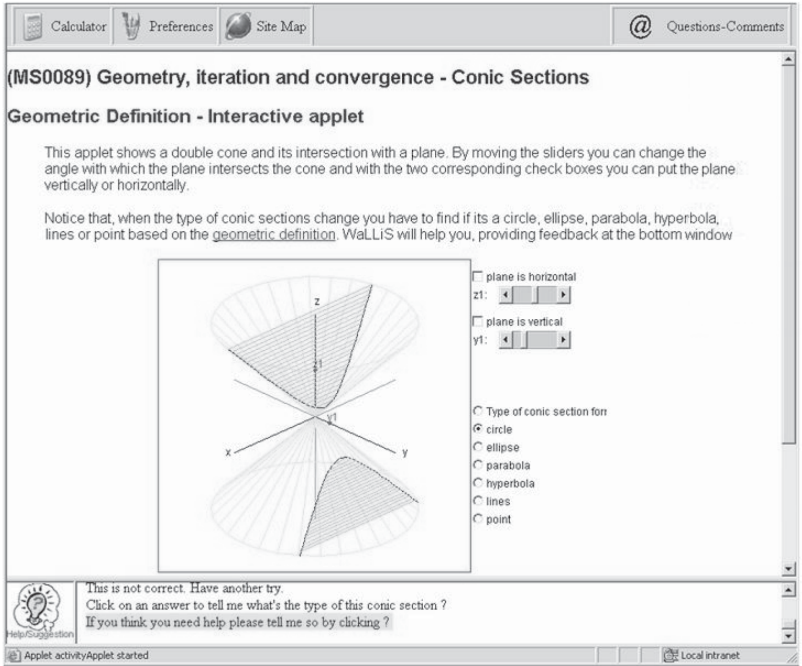


Figure 1: An interactive activity in WALLIS

The feedback is delivered while students interact with multiple-choice questions or other self-practice interactive applets or JavaScript activities (which are authored using appropriate templates). In parts where students need to provide a more explicit answer (for example an integral), this is assessed by Maple on the server side. The exploratory parts are implemented with Java Applets (build with objects that Java Components for Mathematics and DANTE [9] provides) firstly because of their platform independence but especially because of the growing list of freely distributed mathematical applets that can be integrated easily into the system. The in-house built applets communicate further with the feedback frame providing additional help by targeting students' interaction; helping them interpret their actions in a meaningful way (see figure 1). Inspired by previous research in the Artificial Intelligence in Education (AIED) field, and using an enhancement of the feedback mechanism

developed in DANTE [9], the system provides scaffolding feedback and associated hints on their misconceptions derived by rules set in advance.

In addition the system "knows" which page the student is currently accessing and if needed it can provide suggestions on what to do. It will also prompt the student to explore the page completely if they have not done so. The mechanism is still "naïve" but current research aims to enhance it. Ultimately, WALLIS should use AI technology to maintain a cognitive model of the learner that would include aspects such as prior knowledge, abilities, and misconceptions as well as affective characteristics such as motivation, effort, and confidence. These will enable the system to provide suggestions on what students should study and direct the student through an appropriate learning style.

2.2 The system's implementation

During the design and development of all of the system's components, a user-centred methodology was employed which included carefully constructed student questionnaires and data analysis. In particular, observations during students' interactions revealed aspects that would have been neglected otherwise and helped to adjust and improve the system. This iterative "design, use, feedback, re-design" cycle is crucial in any such system. Initially, WALLIS was pilot tested and used by first year science and engineering students only to provide additional support to the other conventional methods of teaching. These students are still using WALLIS to explore materials, interact with the activities and revise for their exams. The system is still limited in content but further materials are being developed especially for this target group (which is our main one due to the many problems they face).

2.3 Assessment in WALLIS

The interactive parts that were briefly described in the previous section could be considered as "low-stake", formative, continuous assessment. Students can interact with them in their own time, receive detailed feedback and learn by experimenting with them. These are not CAA as such. Therefore it was decided to extend WALLIS to include more formally assessed parts. These are described in the following section together with the pilot-test. Following the success that the latter had, and for the reasons mentioned at the introduction, further work will focus on formally incorporating assessment into the system.

3. CAA Pilot-Test

3.1 The assessment question

The long-term goal is not to "re-invent the wheel" but to integrate an already tested and efficient CAA system (such as AIM [10,11,12]) into our package. But to avoid implementation constraints in the short term, it was decided to use a limited but efficient (at least for this pilot-test) JavaBean to assess and mark the students' answers (in a similar way that AIM does so that integration of the two systems may be possible later). One summative question has been designed for a "Geometry, iteration and convergence" course attended by first year honours students. The timing (Easter holidays) permitted the dissemination of an independent subject (conics) entirely on-line. Students had to learn several aspects of conics (such as diagonalisation, standard form, eccentricity etc) from the available materials and then complete a question (inspired by a question from the AIM system). The question comprised several steps which afforded different types of answers (mcq, matrix, numerical, fractions) by asking students to find the associated matrix for a randomised quadratic equation and its eigenvalues, calculate the standard form, classify it, find the associated rotation matrix, the angle of this rotation and its eccentricity (see Figure 2). The question was randomised by fixing the eigenvectors and choosing a range of eigenvalues and constant terms.

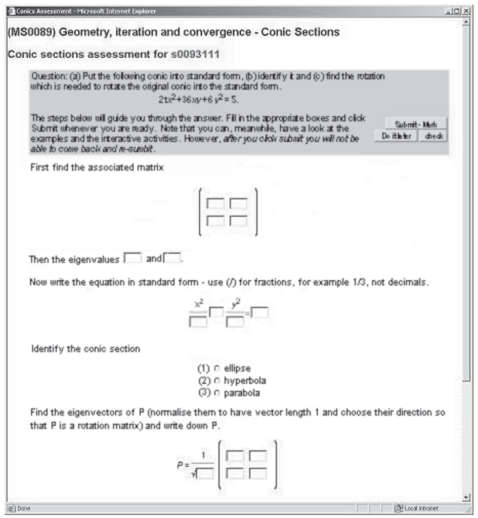


Figure 2: The conics question used for the pilot-test

MAPLE was used to mark the question so that the answers were validated in a semi-automatic way to avoid any surprises in the input, as it was the first time such an approach was being tried. Therefore students did not receive their marks or feedback immediately. Of course, this mimics the real classroom situation. This allowed running some tests before assigning the marks and the marking scheme could be adapted to be fair to everybody taking into account aspects that might have been neglected. In this way, by using a combination of algebraic procedures from MAPLE and the Java Servlet, a variety of robust questions were produced, marked appropriately, with account taken of mistakes that propagate in subsequent parts of the question and produced effective feedback (see section 4).

3.2 Student's overall performance

Most of the 153 students who are actively engaged in the course did take part in the assessment. From the 30 students that did not 12 are at the top 10% of the class. Because there was no other compulsion to complete it and the contribution towards their mark was not so significant they probably did not bother to do the assessment. Another 5 of them were at the bottom 10% of the class and do not attend the lectures (some of these may have already decided to drop the course).

Looking in detail at the results, it was clear that some students would have been unfairly assessed if the system did not give partial credit for certain mistakes that they made in one part which had knock-on effects on other parts. This "injustice" was adjusted by using MAPLE to follow through the student's solution and check, for instance, if the eccentricity value is consistent with the wrong standard form they found in the previous part. This, apart from making the system fairer for the students, helped to provide more appropriate feedback and avoid unnecessary prompts and repetitions (see section 4).

Such simple details (which are often neglected in CAA) biased students against CAA, giving them the feeling that their work will never be as fairly assessed as it would if it were marked by a human (see similar reports in [13). For this test, the overall use of the assessment did not differ from their usual coursework. In fact, their results indicate that even the distribution of their marks was preserved. This can be seen graphically in Figure 3.

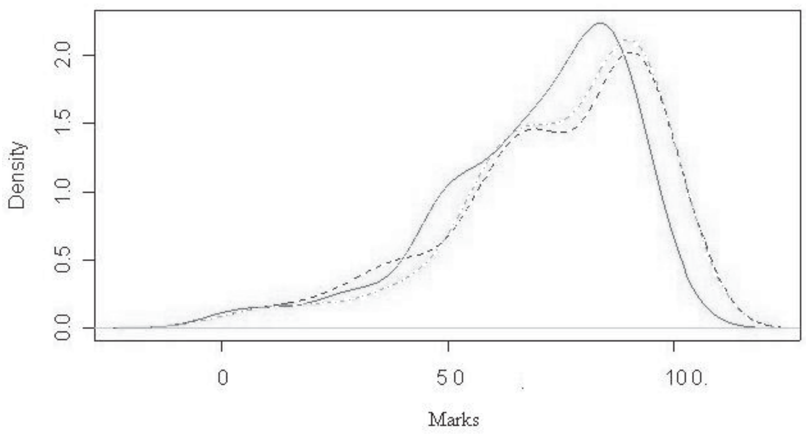


Figure 3: Student's overall performance

The solid line shows the distribution of students' performance in this course (based on previous assessments and small projects). The dashed line shows the distribution of the initial marks the students would get if partial credit had not been incorporated, and the dashed and dotted line shows the adjusted marks. In fact, the results indicate that students performed better in general but this could simply be due to the level and difficulty of the assessment itself (bear in mind that students would have to complete this exact assessment question anyway). There are too many factors to be able to run effective experiments to establish the real reason. Nevertheless, the significance of the above results is even greater when one realises that the concept was not covered by either the lecturer or the tutors and at least indicates that overall students were able to cope with the change in both the delivery and assessment method.

3.3 Individual differences

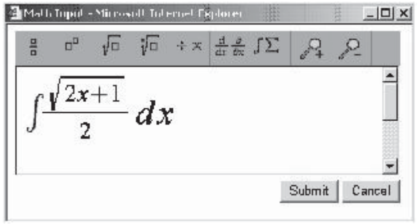
On the other hand, there were 12 students that fell below their previous usual performance. To look for reasons from their relative poor performance all but one were interviewed. From the interviews it was obvious that previous knowledge was the most significant factor and not the concept or the system itself. For this assessment, the bulk of the marks come from finding the standard form and the appropriate rotation matrix which strongly depend from finding the eigenvalues and using them appropriately. During the interviews students commented (and the questions verified) that they had an overall good understanding of the concept but they faced problems with the algebraic manipulation which was implicitly assumed in the related activities and examples. Most of the students said that, for the interactive activities, they were able to follow the examples and reproduce the given method. Unfortunately, some of the examples shared the same numbers which led to students' difficulty in generalising this for the assessment. This shows the importance of carefully thinking through the content that is made available on-line, although it seems clear (and students when asked to comment on this aspect agree) that this issue would not have been better dealt with in lectures. To put it with a student's words "lecturers hardly cover one example and often think we know bits from previous years which I probably forgot". Despite the understandable reasons for this happening (limited time) it is still a problem as far as students are concerned and at least with a system such as WALLIS a student can go through many, different examples and self-practice material.

To evaluate further the results of our pilot-test 10 more students were interviewed following random selection after separating them into different groups who showed significant positive or negative correlation between the amount of interaction with the system, the mark they achieved on the CAA and their previous abilities. While further analysing our data to improve some aspects of the system itself, it is obvious that students' previous knowledge and their ability in algebraic manipulation was the most significant factor in relation to performance. Despite this factor, their overall understanding of the concept was satisfactory. This is evident from several parts of the assessment such as finding the standard form, classifying conics and finding eccentricity especially if algebraic mistakes that occurred are not taken into account. It remains to be seen how these results correlate with the examination marks, which will remove the effects of intermediate steps and test their knowledge better. Finally, most of the students commented favourably on such an approach especially if it used for supporting teaching and delivering formative assessment, various examples and activities and not for substituting lectures or the conventional final assessment.

3.4 Input and other interface problems

It can be strongly argued that the design of computer software is one of the most influential factors for both encouraging use and in the effectiveness of the product. When talking about educational software the stakes are even higher and small problems can have significant but often difficult to detect effects. Therefore, as already mentioned, during the early design and development of all of the system's components, a user-centred methodology was employed which included carefully constructed student questionnaires (with the help of experts in evaluation and HCI fields). In particular, observations during students' interactions and interviews revealed aspects that would have been neglected otherwise and helped to adjust and improve the system.

One of the most important issues was already spotted in the first pilot tests. During their interaction, students were dissatisfied with having to type and understand linearly typed mathematics (see similar problems in [4,13]. Although it could be argued that this is something important that students should learn, it is surely not the right context to do so. The cognitive load to understand and type in this format obstructs their learning of more important aspects at this stage (especially when it comes to quickly interpreting the given feedback).



To allow for a more student friendly notation, WebEQ (www.dessci.com) applets for the input, and MathML (www.w3.org/Math) were employed and the DOM2 interface (www.w3.org/TR/DOM-Level-2-Core) provided the feedback.

Additionally, although a solution cannot be provided, we believe that the way the question is setup often guides the student by limiting the type of answers they can give. This issue has pedagogical significance as understanding the type of answer you are expected to give is an important skill itself and further research should shed light on it.

Similarly, by adopting an iterative design, rather than a more classical software engineering approach, several other problems were located which are not describe here as they relate more to ILEs and their contents rather than CAA itself but instead promote similar methodologies while developing CAA software as they can provide insight on students' needs and lead to systems that students can use effectively instead from struggling to "work around" their requirements.

4. Effective CAA

4.1 Iterative authoring

The iterative design of a system as described above can be extended to question authoring. Our experience, and that of other users of CAA software, shows that questions can rarely be accurately designed the first time: one is bound to

face problems that only occur after the first use. Apart from the importance that this could have in student marking, even slight problems can have significant impact in their learning. When it comes to randomly formed questions there are several things that can go wrong and tools that could help the question designer validate them are needed but until their appearance one has to think very carefully before employing certain types of questions.

4.2 Effective Feedback

Feedback seems to be the most important issue and the strongest asset in computer based learning especially in relation to the problems mentioned in the introduction where traditional ways of assessment and providing feedback are starting to reach their limitations. Tutors, no matter how well trained they are, are bound to face difficulties in locating certain misconceptions and commenting on these, let alone the time constrains that they face. Finding every student's error and writing appropriate and effective comments is often difficult as well as time consuming.

Additionally, commenting on work that was done a week before is not so motivating for students. They often forget the process that they went through in order to produce an answer and a tutor's comment (if ever read) may be worthless at that point. Students usually focus on the marks that they get rather than the actual feedback given and the correction of the misconception. This is when technology applied efficiently can have a significant impact.

Charman and Elmes in [6] have a complete list of recommendations for writing feedback which overlaps with recommendations that teaching centres often suggest to lecturers and tutors (for example see [14,15]). With reference to the feedback's style and tone systems such as WALLIS can provide consider:

- *Make clear if any response is correct or not*

  Adaptive pages can provide visual feedback in a variety of forms including appropriate marking. For WALLIS transcripts ticks and crosses are used to show if something is correct or not and help students focus and correct their mistakes.

- *Give credit for answers that were nearly right and "follow-through" errors to see if subsequent work is accurate*

  This issue relates to partial credit [13]; a major concern in the CAA field but not something that cannot be addressed by technology efficiently. For instance one can exploit the power of computer algebra systems for algebraic manipulation or even theorem provers (such as Omega, Isabelle or Clam) for higher skills such as problem solving. As mentioned before, a much more customised approach was followed (which it is now plan to extend and generalise) which assigned credits for correct parts although previous answers in the same question could have been wrong.

- *Give feedback on correct responses too*

  This seems important, in particular for multiple-choice questions, the answer of which could have been arrived at by chance but also in other questions that a wrong reasoning process would bring the correct result. Even if the reasoning and answer are both correct, feedback can serve its pedagogical purpose.

- *Consider the timing and persistence of feedback*

  There are several issues involved with the timing of feedback. As said before the students were frustrated that some of their interactions cannot be recorded and printed. Comments on similar systems have shown that time limited assessments hinder students' effective use of it. Immediate feedback on each question can be very effective but the lack of time to read and digest this could be an important issue. Students are used to feedback that is preserved in a transcript form after the assessment is over and which they can print and keep in their records for their reference. For these reasons and although students can access their results anytime, these are also in a format that can be easily printed.

- *Consider using graphics (or tailored interactive parts)*

  Graphics or interactive parts can provide more information than a simple explanation and make the feedback much more interesting. Such an approach may be used later and it is worth noting that the available technology exists (for example one can use MAPLE to generate images, as AIM does, or on the fly PHP-constructed gifs) and could surely help automating this process that could be difficult to achieve especially when it comes to randomised questions.

- *Give pointers to further learning opportunities and information targeting the mistakes students made.*

  Once again, adaptive pages can achieve this by suggesting further material

to study, revising certain ones and interacting again with examples. This is one of the next steps in the development of WALLIS and what other systems are trying currently to achieve (such as ActiveMath), where a suggestion mechanism can direct students to appropriate pages.

- *Allow for simple mistakes but penalise and provide guidance for repeated errors.*

  Although this issue can be sometimes difficult to be addressed by computers, technology has the potential to be more effective than tutors. Especially in big tutorial groups, it is difficult to remember who makes what mistakes. If a student repeatedly manifests a misconception then apart from being penalised so as to motivate her to learn she should be also guided to further tasks or reading materials. This is where WALLIS and other intelligent tutoring systems can help by keeping a model of the student.

- *Avoid simply spelling out a full solution*

  This recommendation must be balanced with the previous one. Just providing references to course material or spelling out a complete solution could often result in lowered motivation for the student to work the example again. Usually students know where to find what and repeatedly pointing to the textbook does not improve their learning. Instead it is better to provide an outline of key steps, followed by the final answer.

- *Explain why an answer is wrong and how to derive the correct one (focusing on where student could have gone wrong).*

  The three last recommendations are probably the most important ones. Anyone involved in education knows that this is the essence of tutoring. Research (in [17]) in one-to-one tutoring has shown that this type of tutoring is the most effective of all and one of its important factors is tailored feedback for students' errors. Human tutors can easily locate and "root out" students' mistakes, despite the fact that in certain subjects (such as logic) this is quite difficult. Computers on the other hand may face several other problems and are often limited to simply deciding if a student is correct or not and spelling out the correct answer. Students often consider that as nothing other than another example that they cannot understand. Fortunately, technological advances and research in the field can provide several more elaborate explanations. It is much more effective to explain why the given answer is wrong, target their reasoning process and their mis- (or missing) conception. [4] and others have employed such approaches with AIM (or other systems) to provide feedback on erroneous procedures. For the pilot-test assessment multiple erroneous formulae were given to a simple rule-based JavaBean that could produce the feedback according to the assumed mistake based on the answer given.

Having described the above recommendations and the way that technology can be exploited to achieve them it seems clear that CAA can be more effective than it currently is and systems have not yet showed their full potential.

5. Authoring and Re-Usability

Everyone involved in question and content authoring is aware of how time consuming this process can be. If the complexity of a system is added to this then one is left to struggle with issues that should be minor details and not relate to pedagogy. The problems are twofold. On the one side, there is the reservation to use materials developed by others because of the idiosyncratic ways in which different subjects are taught across institutions. However, projects such as MathWise have clearly demonstrated that this issue is less important when talking about assessment and especially in subjects (such as mathematics for science and engineering) where a more or less national curriculum exists and well-formed collaborations can result in re-usable content. On the other side, even if someone wants to share content, the difficulty of transferring materials across systems and the proprietary format each one uses, makes authoring so time consuming that many people decide to re-develop than reuse. The sharing business becomes more difficult when universities are already using mainstream educational software and the low budget of a department could forbid adopting other systems just to use materials that colleagues have already authored. Even if one adopts open-source or community-developed software (minimising the cost), the overhead of learning how to author for them and the need to use previously developed material is so big that people are against trying it.

Moreover, authoring materials in proprietary and difficult to change or share formats (despite the advantages that these formats seem to have) can only provide a short-term and context-dependent solution that cannot be neither reused nor adapted by others. Even if one neglects the time, resources and funding misuse that such a solution implies, one should also bear in mind that any educational content is bound to need revision and changes either because of instructors' or, more importantly, learners' needs. Developments often follow the same or similar materials, activities and questions, re-inventing the wheel just because of the lack of collaboration and the difficulties of transferring content between systems.

On the other hand, bodies such as the IMS, IEEE, ISO and other initiatives (such as the European CEN/ISSS, Prometheus and CETIS) have the potential to provide valuable solutions, which (if produced by generally accepted requirements) could result in content which can work with different applications, increase the range of materials and questions available to everyone and help institutions use systems that have the features they want, reusing their previously developed materials, without worrying about integration, data format and other technical issues. This (despite the fact that it often sounds like nothing more than a dream) can only be achieved by well-formed interdisciplinary collaborations and feedback from all interested parties. Research in relevant and more appropriate fields, such as the emerging Mathematical Knowledge Management (MKM [18]) field, look very promising. The field has a significant overlap with the mathematics education field and could surely need the feedback and specification requirements that mathematics educators could provide. Classifications and approaches such as the ones examined in the previous maths-caa article [19] combined with requirements such as the feedback types discussed above or question types others have described [for example 5] in conjunction with elaborate techniques and specifications (such as OpenMath, OMDOC and ActiveMath [for example 20]) may have a higher initial authoring cost but the further development of tools will soon have a significant impact on our field and ease storage and content/question authoring.

6. Future work

Recognising that the worst problem in component re-usability is that most systems follow customised and not a re-usable approach specifications and standards were followed closely. It is the intention to release some of the components of WALLIS so that they can be re-used. Future work will focus on releasing the adaptive tree-like navigation map as well as any student management system that emerges from the work. Similarly, while developing further content and questions for WALLIS (and probably AIM) it would be good to be able to share those with others. So far no specific standard has been followed (because there was little material developed), but it might be sensible to commit to any reasonable emerging one.

Finally, it would be appropriate not to lose sight of the target audience. Mathematics departments across the UK are having to cope with an increasingly heterogeneous cohort of students with diverse abilities, needs and background (and this is true even when they have apparently taken the same school examinations). CAA has a vital role to play in tackling this problem. Unless staff and tutor numbers can be increased auxiliary support for such students will have to be provided. One argument against this is that tackling the needs of, for example, students from disadvantaged backgrounds cannot be reasonably done with CAA because such students have had no experience of computers. This is becoming much less of a problem. The drive to equip all schools with a basic IT provision is beginning to be realised and it is rare to see a student from any background who has not had some experience of computers. Moreover, there is a greater desire on the part of these students to use IT systems perhaps because they have more to gain from them. This target audience should provide a boost for our industry because there is real pressure to deal with their special needs.

References

[1] Kent, P. 2002. Should we reassess the role of Mathematics in Engineering education? Available online at: http://mathstore.ac.uk/articles/maths-caa-series/dec2002/index.shtml.

[2] London Mathematical Society. 1995. *Tackling the Mathematics Problem* - a joint report of the LMS, IMA and RSS. Also available online at: http://www.lms.ac.uk/policy.

[3] Mavrikis, M. and Maciocia, M. 2002. Developing WALLIS; a web-based system to enhance mathematics teaching. ICTM, Crete, Greece (also available online http://www.maths.ed.ac.uk/~wallis/)

[4] Pitcher, N. 2002. Assessment with Mathwise – Lessons, Limitations and Learning Curves. Available online at: http://mathstore.ac.uk/articles/maths-caa-series/sep2002/index.shtml.

[5] Sangwin, C. 2002. New opportunities for encouraging higher level mathematical learning by creative use of emerging computer aided assessment. Available online at: http://mathstore.ac.uk/articles/maths-caa-series/may2002/index.shtml.

[6] Charman, D. and Elmes, A. 1998. *Computer based assessment Vol 1, 2: Case studies in science and computing*, SEED Publications, University of Plymouth.

[7] Crawford, K., Gordon, J. and Prosser, M. 1998. Qualitatively different experiences of learning mathematics at university. *Learning and Instruction*, **8**(5) 455-68.

[8] Woolf, B. P., Beck, J. E., Eliot, C. and Stern M. K. 2001. Growth and maturity of ITS. In Forbus, K. D., Feltovich, P. J. and Canas, E. (Eds). *Smart machines in education*. AAAI/MIT Press.

[9] Mavrikis, M. 2002. *Towards More Intelligent and Educational DGEs*. Master's thesis, The University of Edinburgh, School of Informatics; Artificial Intelligence.

[10] Klai, S., Kolokolnikov, T., Van den Bergh, N. 2000. Using Maple and the web to grade mathematics tests, *International Workshop on Advanced Learning Technologies*, Palmerston North, New Zealand.

[11] AIM has been recently released under a GNU license see http://sourceforge.net/projects/aimmath.

[12] Kyle, J., Sangwin C. 2002. AIM - a parable of dissemination ICTM, Crete, Greece.

[13] Beevers, C. E, Youngson, M. A., McGuire G. R., Wild D. G. and Fiddes, D. J. 1999. Issues of Partial Credit in Mathematical Assessment by Computer , *Alt-J* **7**, 26-32. Also available online at: http://www.calm.hw.ac.uk/publications.html.

[14] Arthur, D. 1995. Problem Solving Classes, in [16].

[15] Anderson, C. 1995. Student Learning, in [16].

[16] The Centre for Teaching, Learning and Assessment. *Tutoring and Demonstrating*, The University of Edinburgh (in association with UCoSDA). Also available online at: http://www.tla.ed.ac.uk/publications/tutdem.html.

[17] Bloom, B. S. 1984. The Two Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, **13**, 4-16.

[18] Asperti, A., Buchberger, B. and Davenport, J. H. (Eds.) 2003. *Proceedings of the Mathematical Knowledge Management Second International Conference, MKM 2003,* Bertinoro, Italy. Lecture Notes in Computer Science 2594 Springer, ISBN 3-540-00568-4.

[19] Williams, J. S., Barry, M. 2003. TAL - A National Database of Questions - Classification is the Key. Available online at: http://mathstore.ac.uk/articles/maths-caa-series/may2003/index.shtml.

[20] Goguadze, G., Melis, E., Ullrich, C. and Cairns, P. A. Problems and Solutions for Markup for Mathematical Examples and Exercises, in [18] 80-92.