

**A review of factors influencing the dissemination of the London Agreed Protocol for Teaching (LAPT) -- a confidence based marking system.**

**Dr J. A. Longstaffe and  
Prof JWB Bradfield**

**May 2005  
V-18.1**

<b>Table of Contents</b>	<b>Page</b>
<b>Summary</b>	<b>2</b>
<b>Introduction</b>	<b>3</b>
<b>Current use/dissemination of LAPT</b>	<b>4</b>
<b>Potential for dissemination of LAPT</b>	<b>5</b>
<b>Potential obstacles to dissemination of LAPT</b>	<b>8</b>
<b>Recommendations and suggestions</b>	<b>11</b>
<b>Further evaluation</b>	<b>14</b>
<b>Conclusions</b>	<b>14</b>
<b>References</b>	<b>14</b>
<b>Appendix 1 -- Interview Technique</b>	<b>15</b>
<b>Appendix 2 -- Sample question file used by LAPT -- appearance to author and user.</b>	<b>16</b>
<b>Appendix 3 -- Discussion of gender bias in multiple-choice questions</b>	<b>18</b>

## 1. Summary

The authors, who have an extensive background of teaching and the learning in veterinary and medical education have performed a study of the factors influencing the dissemination of the principle of Confidence Based Marking (CBM) as embodied in the London Agreed Protocol for Teaching (LAPT).

LAPT is a version of CBM in which students are asked to assess the level of confidence with which they are answering objective questions. A scoring system rewards high confidence of correct answers and penalises high confidence on incorrect answers.

CBM (LAPT) is currently extensively used within UCL, especially within medical science. Dissemination to other disciplines and other institutions has been facilitated by the development of a Web-based authoring tool and a support website.

This study involved the reaction of the authors to the concept, the website and the authoring procedure (authoring tool, manual and help files), together with informal interviews of two teachers who use the system extensively, four teachers who have expressed interest but so far have not adopted the system in their own institutions, two teachers to whom the concept of CBM were new and lastly a clinical educationalist.

### The study noted that:

- LAPT is currently in extensive use within UCL or UCL related institutions.
- With a few exceptions, although the concept was found to be exciting, the use of LAPT has not by and large spread to other institutions.

### Positive findings include:

- The concept of confidence based marking (CBM) is attractive to teachers in a wide range of institutions.
- Students like it, or at worst are no more negative about it than other forms of objective testing.
- Students experience this as a fairer marking system than standard negative marking.
- There is sound academic proof of concept (<http://www.ucl.ac.uk/%7Eucgbarg/pubteach.htm>).
- Demonstration and personal contact are the best ways of disseminating the idea.
- The LAPT website has been used extensively by teachers within UCL and a few other institutions with no serious problems.
- The LAPT developer is willing to insert questions into LAPT on request.
- Support systems include a comment facility and e-mail support.

### Possible obstacles to dissemination include:

- Sometimes confusing web site design and navigation.
- Explanations of LAPT which need to be clearer and more succinct
- Those concerned with high stakes summative examination have reservations about its adoption.
- Access to question banks containing some poorly written questions, which distract from a consideration of the CBM system
- Pedagogic negativity in the community concerning some forms of objective testing and negative marking as well as fear of gender bias.
- Personal and institutional inertia due to the work involved in transferring to a new system which is competing with other systems already adopted by other institutions.

### Recommendations and suggestions include:

- Continuation of the programme of seminars, demonstrations and personal contact which clearly works to disseminate the concept.
- Redesign of the website
- To make it easier to understand and to separate it from internal student and staff relevant pages.
- To include a succinct and clear "sales" component that answers the teacher question "why should I use this?"
- Giving prominence to a limited number of well written questions which:
- demonstrate the need for knowledge of confidence levels in real life situations
- demonstrate that objective testing can be used for more than the testing of superficial knowledge.
- Possible renaming/rebranding of LAPT.
- Research into the history of the negativity surrounding objective testing and negative marking in order to allay fears which may not be based in reality.
- Use with caution any question types disapproved of by the clinical medical community.
- Making it clear to what extent LAPT conforms to IMS-QTI standards.
- Clarifying editing and question insertion procedures as well as authoring and importation procedures.
- The development of partnerships (some already exist) with commercial organisations, marketing assessment systems and also with institutions like the Open University.
- Targeting dissemination particularly to subject areas where a knowledge of confidence is a clear real-life advantage.
- Avoiding (for the time being) high stakes summative areas until formative use is well established.

### Appendices include

1. Interview technique.
2. An example of a LAPT question file.
3. A discussion of gender bias.

## 2. Introduction

### 2.1 Study Remit

In March 2005 the authors of this report were asked by Professor A. R. Gardner-Medwin to evaluate the LAPT system in terms of its potential for dissemination. The authors agreed with Professor Gardner-Medwin a study which would focus on factors influencing the dissemination of LAPT as representative of Computer-Based Marking (CBM) rather than an evaluation of the intrinsic pedagogic value of the system. It would be written from the point of view of potential adopters -- teachers rather than educationalists (the authors are experienced medical/veterinary teachers both of whom have long-term experience with e-learning). As a result of this small-scale survey recommendations would be suggested for possible improvements and changes.

### 2.2 What is LAPT?

The London Agreed Protocol for Teaching is a computer-based system of objective testing which uses confidence based marking (CBM). This requires students to answer questions at one of three levels of confidence, low, mid or high. If the answer is correct, higher levels of confidence are rewarded by more marks but if the answer is incorrect, higher levels of confidence are penalised by successive levels of negative marking. Claims for CBM over other marking systems include:

- students give more thought to their answers and as a result:
  - are challenged to understand rather than just react
  - subsequent study is more discriminatory
- students with low confidence who know the answers are given a more accurate picture of how confident they should be
- students are trained to assess their levels of confidence accurately -- a life skill which pervades all disciplines but which can be a matter of life and death in disciplines such as Medicine and Engineering
- the system is fairer because it rewards careful thought rather than guessing
- the system requires fewer questions for an accurate assessment of ability.

### 2.3 What the authors have done

1. Examined the LAPT website in detail.
2. Read (one of us only) the supporting literature published on the site.
3. Downloaded material from the site and spent time with questions and authoring
4. Interviewed selected individuals by telephone or face-to-face:
  - 2 teachers naive to the concept of CBM (selected from our own colleagues)
  - 4 teachers who have examined the concept but have not yet adopted CBM (3 suggested by TGM and 1 selected from our own colleagues)
  - 2 teachers who are using LAPT with their students on a regular basis (suggested by TGM)
  - a senior clinical educationalist.

The interviewees had much to contribute and a total of 5.5 hours were spent in interviews by one of the authors. See **Appendix 1** for details of the approach to interviews.

### 2.4 Acknowledgements

The authors would like to acknowledge the enthusiasm and depth with which all interviewees were willing to engage -- as a result, this report contains much more useful information than could have been anticipated.

### 2.5 Caveat

This report is necessarily based on the subjective opinions of a relatively few people and as such may not be representative of a larger group. Nevertheless, the authors consider that the responses obtained in interview were thoughtful and show general agreement.

### **3. Current use/dissemination of LAPT**

#### **3.1 How is LAPT disseminated?**

Over the past few years attempts have been made to disseminate LAPT using a number of strategies. These have included:

- personal and informal support and contact with immediate colleagues
- formal contact with heads of departments of physiology
- demonstrations and posters at conferences and seminars
- encouragement for students to share the system with colleagues in other medical schools
- the creation of a web based authoring system LAPT-lite which is accessible from anywhere
- the negotiation of partnerships with commercial OMR (Optical Mark Reading) products such as Speedwell Computing
- the adaptation of LAPT to fit with at least one Virtual Learning Environment (VLE)
- the adoption of IMS-QTI standards (in progress).

#### **3.2 What is the current dissemination position?**

Information for this section has been taken from Professor Gardner Medwin's report to FDTL so is only briefly summarised here.

It is apparent that LAPT is extensively used within UCL especially within the medical sciences. It is used both formatively and summatively in the preclinical disciplines and formatively in the clinical disciplines. Submissions to the Web-based questions represented some 840,000 Q-A interactions and 5000 hours of student time from October 2003 to November 2004.

An online practice site using CBM in preparation for the paper-based (non confidence based) Bio-Medical Admissions Test (BMAT) resulted in some 32,000 Q-A interactions by school pupils.

As a result of contact between medical students, there has been sporadic use of the system from students in other medical schools nationally and internationally (300 accessions).

LAPT has been adapted to fit with at least one VLE (WebCT operating within UCL and Imperial College London (ICL)). LAPT is in regular use in both institutions, and is also used by Winchester College along with its own proprietary VLE. .

With the above exceptions, although interest in LAPT has been high, there has been little actual adoption of the system by other institutions.

## **4. Potential for dissemination of LAPT**

### **4.1 It has impressed us that the concept of Confidence Based Marking is attractive to teachers**

All of the interviewees find the pedagogic concept attractive. One reaction is worth quoting -- "a fantastic system and he has validated it perfectly."

One articulated the reason for its value as follows -- "... it is necessary for the student to take two decisions, firstly whether the option is correct or incorrect and secondly how confident they are about this, they are thus giving more thought to the answers rather than just guessing."

It is significant that three of the four non-adopters interviewed had been so convinced of the value of CBM as to attempt to disseminate the idea to their colleagues. Reasons that this has so far not worked are discussed below. An excerpt from a staff critique at UCL reads:

*"The introduction of confidence based marking has stimulated a new introspective analysis of the way students revise. Since LAPT penalises overconfident mistakes, it provides a stern wake-up call and stimulates true understanding of concepts and principles."*

One of the naive group (engineering maths) was also so enthused as to present the ideas to his colleagues at a staff away day -- it was received positively and the staff member intends to pursue this further.

### **4.2 Students appear to like it**

Although students were not interviewed directly, published evaluation responses provided by Professor Gardner Medwin seemed to indicate general acceptance and often positivity by students. This impression was confirmed by the other two adopters interviewed.

One quote from a student who had been encouraged to reflect on how computer-based assessment and confidence based marking might have affected learning at school runs as follows:

*"Introducing confidence based marking would certainly make me more aware of the answer I give as I would put more effort into my thinking process."*

Although one interviewee felt that the students would fuss at first, he also felt reassured by the supporting documentation which seemed to confirm that students do come to live with it.

### **4.3 There is apparently sound academic proof of concept**

The author (JAL) who read the academic papers supporting the LAPT system felt that Professor Gardner Medwin had provided a sound academic basis for the basic assertions made about LAPT and this opinion was echoed by several interviewees. Adopters were satisfied that the system was sound and even non-adopters did not challenge the work published by Professor Gardner Medwin.

### **4.4 Use of personal contact, seminars and demonstrations**

It seems clear, not only from the actual distribution and current use of LAPT but also from conversations with teachers who are enthusiastic about the system but have not yet adopted it that personal contact, seminars and demonstrations are much the best routes for dissemination of the concept. Although there are opportunities to improve the website (see below), it is possible that this will function largely as a support tool once teachers are convinced of the initial value of the system. This initial "selling" of the LAPT system has clearly worked best through personal contact and the enthusiasm of its developer.

### **4.5 Use of LAPT website**

It is noteworthy that there has been a markedly different reaction to the website from people who were attracted to the idea of CBM and wanted to use it compared to people who accessed the website to find out about LAPT and CBM. The first group were relatively unfazed by the website design and were able to find their way through to the authoring tools and the specimen questions. The second group were more critical of the website design and found it more difficult to use it to get a sound idea of what LAPT and CBM were about (see below). From this the authors have concluded that the website and the authoring systems are useful tools for those already sold on the concept but less effective as an initial sales tool.

The website does have facilities for the feedback of comments, offers help with technical problems and provides a service for the insertion of questions into the system as well as the adaptation of existing questions.

#### 4.6 Authoring questions in LAPT -- use of LAPT-lite (the web-based software)

##### Pros:

- The only current system which enables the use of CBM in a soundly researched format.
- A workmanlike authoring system -- does what it says on the box.
- Covers a reasonable range of question types and options.
- Gives good flexibility for feedback.
- Allows the incorporation of images.
- Option to author at two levels -- an authoring tool or an editable text file
- Incorporates font formatting options such as bold italics, superscript, subscript.
- Some questions from Web CT can be imported via a conversion tool.
- Is moderately easy to learn.
- The format of the files used by LAPT is simple and can be easily edited in a text format (see **Appendix 3** for an example).

##### Cons:

- Once a file has been created in the authoring tool, editing needs to take place at the text file level.
- The manual supports the specifications of the final text file level rather than the authoring tool.
- It lacks a coherent help/troubleshooting system (although there are help documents).
- The manual and authoring tool present as work in progress rather than a finished system.
- The QTI conversion tool is still under development.
- At the simplest user level, the screen design and appearance are inflexible (although users familiar with HTML can adjust screen design and appearance).

##### Introduction

Only one of us went into this authoring system in depth (JAL) and what follows is a personal response based on a maximum of four hours experience of the system. There may well be things which I missed but I decided that someone who is enthused about the concept, is not going to give up much more than one evening to see whether the authoring system is a practical tool. It has to sell itself in a relatively short time.

##### The authoring system

Authoring is possible at one of two levels -- an authoring tool (described below) or the writing of a fairly readable final text file (technically a JavaScript file) specified in the manual provided.

Within the **authoring tool**, the author is guided through a series of screens with reasonable explanations which prompt the filling in of on-screen forms for the various portions of the questions. It is easy to create questions by cutting and pasting from ordinary text files and the resulting questions seem to work. Interestingly enough, my voice recognition system didn't work with this system -- forcing me to dictate into Word and then cut and paste. I am aware that I am still in the minority in using this particular form of input but if this was easy to correct, it might help.

Questions on offer include:

- True/False
- Multiple Response
- Single Best Answer
- Extended Matching
- Text Entry
- Numerical Answers (including ranges)

One of the strengths of this authoring system is the flexibility of the various feedback options. They provide a good range of ways to respond, to give information and to give different responses depending on whether the student answers correctly or incorrectly.

I was glad to be offered text formatting options such as bold, italic and, importantly, subscript and superscript. The authoring tool also offers a full range of symbols. I found the handling in the authoring tool of the HTML "tag" system for creating the font and text formatting slightly clumsy, especially when I tried to format text retrospectively.

Largely because of the flexibility of the various feedback options, some of the authoring screens proved a little difficult to grasp at first and it would have been helpful to have a tutorial to guide me through step by step with comments and with the forms filled in for me. It would also have been useful when I was in an individual screen to be able to click on a "Help" button which explained what this particular bit did -- maybe giving a picture of the finished question as viewed by a student with the relevant part highlighted.

It is not possible to return to the "fill in the box" authoring level once a file has been produced. Being able to step back through the authoring tool would also be an advantage. But this said, the format of an exercise file for the common formats is quite straightforward.

### **Downloading LAPT-lite**

Downloading for off-line authoring was easy and the 5.5 MB zip file came down in just over a minute on my broadband connection. Unzipping and following the readme instructions was correspondingly easy. The download is intended for offline use, and gives the same functionality as the webpages. This is mainly provided as a facility for students at present, rather than staff. Triggering the appropriate .htm file produces a copy of the same rather offputting LAPT-lite page that appears on the UCL website. Various helpful documents eventually show up but if they could be placed under a menu which says "How to Use LAPT-lite" it would be helpful. This would be much better as a menu driven series of pages. It should be said that is not normally necessary to download LAPT-lite unless the author needs to work offline.

Instructions recommend running a newly authored file by accessing the primary source of the software at UCL on the web, in conjunction with the local exercise file. I was confused about the necessity for this, though I gather it is equally possible to use the downloaded software from a local drive. Following a conversation with the developer I understand that the **LAPT** system can be used with exercise files placed on a server elsewhere, either using the system program files accessed from UCL, or with copies saved and maintained elsewhere. This is a strategy to avoid having to update the software on multiple remote sites unless remote users wish to mount and maintain the software themselves. This does need a clearer explanation in the support documentation, and is now set out in a document for teachers (<http://www.ucl.ac.uk/lapt/cbmflyer.doc>).

### **The manual**

The current manual which guides the author through the basic text file is essential for editing because of the fact that the authoring tool is a one-way process. It gave me what I needed to edit the files I had produced already. It is really more of a specification of the syntax than a manual. A manual written at the level of "1.Do this..., 2.Then do that..." would be helpful but not essential. I was able to experiment with what the various bits did and understand the concepts fairly rapidly.

### **Conclusion (on the authoring system)**

**LAPT-lite** is a reasonably sound workmanlike authoring system of which the chief advantage is its support for CBM.

The authoring process is relatively easy to learn and enabled me to produce coherent questions within an acceptably short time.

The front end of the authoring process makes it clear that this system is still under development -- this might be off putting to potential institutional adopters. I wonder if it would be possible to mount a version which had been thoroughly tested in specific browsers

Although there is comprehensive support available by e-mail from the developer, it needs to be clearer that this is so. Attention paid to the organisation of a visually attractive and coherent authoring support system would go far to motivate remote users.

In conclusion, it shouldn't be too difficult to make this system clearer and more attractive to potential new authors. This is certainly worth doing in view of the value of the basic concept.

## 5. Potential obstacles to dissemination of LAPT

### 5.1 Website

Both the authors and some of the interviewees had some reservations and suggestions about improvements that could be made to the website to make it a more effective tool for dissemination. There was less criticism of the website as a tool for the authoring and delivery of questions.

#### 5.1.1 Web site design

Reactions to the website as a dissemination tool were mixed. Reactions included the following:

- the website was too complicated
- links and navigation were confusing
- there were no clear hierarchies of information.
- The home page for the **LAPT-lite** system was particularly confusing and users found the list of text starting with "PROBLEMS..." off putting. They felt that this complex information for regular users should be buried at least one layer down with clear links.
- In at least two cases, new users had not even been aware that there was an authoring system or tool available for their use.

In contrast, other individuals had fewer problems with website design and appeared to find what they needed. Those who had been familiar with earlier versions of the website said it had been "much improved".

#### 5.1.2 Description and explanation of LAPT -- the "sales pitch".

Naive users (including one of the authors) felt that the explanations as to why LAPT was a good idea were difficult to understand. The following points may be useful.

- The introduction of statistical graphs at the initial explanation stage is confusing to the non-statistically numerate.
- Some users were initially confused by the concept of "confidence score".
- The approach on the homepage is basically an explanation to students as to why LAPT is a good idea -- there needs to be an explanation aimed at staff.
- The approach to the students mixed up the "why use LAPT" and "how to use LAPT" points.
- Some of the assertions as to why LAPT was a good thing were challenged as being "not necessarily so" (see under recommendations 6.3.1 -- pp 11-12).

The authors and interviewees felt that there was a gap between the obvious and intuitive fact that CBM should be an advantage and the equally valuable and detailed academic "proof of concept" papers. They felt that there was a need for a middle position which provided an easily digestible series of statements which were clear and could be backed up by the academic documents if someone wanted to go further.

#### 5.1.3 Authoring support

Please see the section on LAPT-lite above.

#### 5.1.4 Mixed question quality

Some users were particularly critical of the fact that when they tried to sample from the large numbers of questions available on the site, many were felt to be poorly written (this was also evident in one of the student critiques read). They also made it clear that this was no different to the situations in their own departments but pointed out that the poor questions distracted from the consideration of CBM as a concept. It was suggested by several interviewees that it would be a good strategy to provide initial access to a relatively few well-written questions (see recommendations below).

#### 5.1.5 Relevance of confidence

Interviewees pointed out that, although accurate assessment of confidence is a universally applicable life skill, there might be a difference between disciplines of its perceived advantage. It was suggested that examples be given from areas such as medicine where it is absolutely clear that the doctor needs to know how confident he or she is when making decisions which involve patients. This sort of example would also influence the attitude of students to CBM – the greatest fear of most medical students and junior doctors is that they are going to kill someone through getting it wrong. Being able to judge levels of confidence would be a great motivator for these students. Maths, languages and law also stand out as disciplines where students lose out by making decisions or jumping to conclusions based on insufficient reflection.

### 5.2 Pedagogic negativity

A number of interviewees had come across prejudice against objective testing and negative marking when trying to persuade their colleagues of the advantages of CBM. Additionally, some colleagues felt that (or had heard that) CBM discriminated between genders.

### 5.2.1 Resistance to objective questions

Only one of the interviewees had reservations about the use of objective questioning (MCQ and T/F) but most had come across objections from colleagues which in many cases did not seem to be supported by reference or evidence. There did however appear to be some difference of opinion between educationalists some of whom are against this form of assessment and teachers who are often attracted by it.

One major objection seems to be a generally held view that objective questioning only tests superficial knowledge. It was suggested that the website could give examples where objective testing is used to test much deeper and complex learning (see recommendations below). There are in fact a number of "thinking skills" type questions available in the Bio-Medical Admissions Test (BMAT) section of the website but these need to be given more prominence (or even extracted and used in a limited group of illustrative questions -- see recommendations below).

In a helpful interview with a clinical educationalist one of the authors was able to get a much more focused picture of these reservations. In short, in the clinical medical area, and not simply limited to UCL, there is specific discomfort with the Multiple True False Question. The perception seems to be that in order to be highly discriminatory, these questions can tend to test smaller and smaller items of factual knowledge which are often of low relevance/significance -- this is against the clinical teaching philosophy of encouraging analysis and problem solving. The most important finding in this interview was that these objections do not appear to extend to Single Best Answer questions nor do they extend to Extended Matching Questions. The interviewee was also extremely supportive of the basic principles behind LAPT.

The evidence behind the above objections appears to be relatively old and based on work done in secondary schools in the US. Although the precise nature of the evidence was not clear to the authors, what is clear is that there are very strongly held opinions across the clinical medical area (including the GMC and the Royal College of Physicians). As a result, whether the objections are evidence based or not, the disputed question format needs to be used with care.

More generally in higher education folk myth, the whole of objective questioning seems to have fallen into disfavour. This problem could well be addressed by giving good examples of questions which address deeper learning objectives than superficial factual knowledge. However it should also be pointed out that one interviewee uses objective questioning precisely because he needs to know the students' level of knowledge of basic facts early in a course. It is also interesting to note that this general background opinion seems to be changing as universities become more committed to (and experienced with) Computer-Based Assessment (CBA) -- See "Cultural timing" below.

### 5.2.2 Resistance to negative marking

This was interesting. Different teachers appear to have very different experience of the attitude of students to negative marking. One of the authors of this report has the experience that students really dislike negative marking and feels that this might prejudice them against the use of CBM (where negative marking seems at first to be even more pronounced). On the other hand, other teachers, including those currently using LAPT do not experience the students as being particularly resistant to negative marking. One of the "naive" interviewees said "my goodness, this will cause a fuss with the students!" But he then went on to say that he felt this issue had been adequately tackled in the supporting papers which confirm that students do come to live with it.

The clinical educator felt that it should be possible to reward accurate assessment of confidence and punish inaccurate assessment of confidence without a system which technically involved negative marking. One of the other interviewees (engineering maths) also felt that he would want to develop a marking system which still discriminated between the two but was not negative.

Importantly, the developer has pointed out that LAPT/CBM is a variable system of negative marking where students do appreciate the option of being able to avoid negative marking by expressing low confidence.

### 5.2.3 Possible gender bias

Several interviewees felt uncomfortable about using CBM as they had been told that there was a risk of gender bias. It seems to the authors that the absence of demonstrable gender difference at UCL using LAPT is convincing (*Gardner Medwin & Gahan 2003*). The authors have included a discussion of possible gender bias in Appendix 2.

Given the different opinions around objective questioning, negative marking and gender bias, it might be useful to perform further research into the literature around these issues in order to stimulate informed debate and (possibly) reduce prejudicial attitudes (see recommendations).

#### 5.2.4 Cultural timing

Several interviewees felt that there were cultural changes currently taking place which might make objective testing and negative marking more acceptable. Five years ago the use of objective testing in UK higher education was largely paper-based and was generally unpopular especially with educationalists. Recent pressures on staff time together with the increasing availability and sophistication of Computer-Based Assessment (CBA) have motivated many staff and institutions to revisit the use of electronic assessment and objective testing. Interviewees gave the impression that resistance to these concepts has been reducing over the last few years. On an organizational note, at least one university (Bristol) now has an identifiable committee whose task it is to examine CBA options and make appropriate decisions. The emergence of such identifiable bodies should also be a positive factor for potential dissemination of CBM.

#### 5.2.5 Formative versus Summative

A number of the interviewees were more comfortable with the use of the LAPT system for formative rather than summative testing. Most of their discomfort around summative testing was the lack of clarity around where to set the pass mark. One teacher said -- with the current system, a student that gets everything right at the lowest confidence level only gets 33% -- should they be failed? Although it is clear that summative testing is successfully used in preclinical disciplines at UCL, the authors feel that there needs to be clear guidance on pass marks before new users will be confident enough to use the system for summative testing. Although experience at UCL shows that the kind of specific concern raised above would seldom if ever occur with practised students, institutions need experience with formative assessment to allay concern.

In the experience of one of the authors (JWBB), unless objective questions are embedded in instructional material, students want to practice the form of assessment that they can expect in the summative examination. In practice, one of the pressures to use CBM in summative assessment at UCL was the student preference for this as a fairer marking system.

### 5.3 Personal and institutional inertia

It probably goes without saying that change is difficult in the current academic atmosphere. Teachers are increasingly busy and the idea of instituting a new assessment system with all the work involved in the transfer of questions from another system can seem very daunting.

Institutions tend to support specific authoring/assessment packages -- the difficulty lies in selling the concept of LAPT to the individuals or committees that might authorise institutional support. The suspicion of several interviewees has been that the pedagogic objections discussed above may have been used to disguise the fact that there was simply not enough time or motivation to institute change.

For potential adopters, it has often been difficult to identify the decision-making process that needs to precede the adoption of a new assessment system by a department or institution. It is apparent that, as institutions embrace computer-based assessment in a more organised manner, they are creating bodies whose responsibility it is to recommend good practice. As these decision-making processes become clearer, it could become easier to encourage the adoption of CBM. Timing is also important -- if an institution has just changed or adopted another online assessment system, it would not be a criticism of LAPT or CBM if they failed to adopt it.

One significant factor cited by more than one interviewee was the continuing bias in academic circles towards research. This makes investment in improving teaching a relatively low priority in many areas.

Many departments and teachers are currently wrestling with issues such as:

- the production of good questions
- the structure of fair marking systems
- the delivery of questions over secure systems

... and these take priority over the adoption of CBM which is seen as the icing rather than the cake itself.

Interviewees have suggested some possible approaches (see under recommendations below).

## 6. Recommendations and suggestions

The adoption of these recommendations and suggestions will of course be a function of available resource in terms of time and funding. Although items have been divided into recommendations and suggestions, it should be left to the discretion of the LAPT developers as to which ones are adopted and how they are prioritised.

### 6.1 Seminars and demonstrations

#### Recommendation 6.1.1

*Continue a programme of seminars and demonstrations. This is clearly the most successful method for dissemination of the initial concept.*

#### Recommendation 6.1.2

*What also appears to have helped is ongoing personal contact and support. This seems to be the reason for the extensive dissemination within a relatively few departments and institutions. One way of extending this would be to provide support for "champions" situated in other institutions who could then disseminate and support their colleagues. As yet unsuccessful attempts have been made by enthusiasts to disseminate at the University of Bristol and in the Open University and an examination of the ways in which these enthusiasts could be supported might bear fruit.*

#### Suggestion 6.1.1

*One interviewee has suggested that seminars be as "hands on" as possible giving participants a chance to focus on and solve some of their own problems rather than just watching a demonstration.*

### 6.2 Website design

#### Recommendation 6.2.1

*Redesign the LAPT website with the following characteristics:*

- *A less crowded appearance and with clear navigation*
- *A menu system with hierarchically arranged information and links*
- *A home page directing users according to need -- for example:*
  - *"I am a teacher and want a brief explanation of LAPT and its advantages."*
  - *"I am a teacher and want to access the underlying academic support for the claims of LAPT. "*
  - *"I am a student and want to know why LAPT is being used/introduced. "*
  - *"I am a student and want to know how to use LAPT. "*
  - *"I am a teacher and want to look at some specimen questions."*
  - *"I am a teacher and want to try to make some questions of my own."*
  - *Etc*
- *It would also be useful to provide a link to a page of "Common Misconceptions" in which the gender issue could be clarified and issues like objective testing and negative marking discussed with appropriate links to further reading -- see **Appendix 3** for a start.*
- *The **LAPT-lite** page should be clearly labelled as an authoring system with its manual.*
- *Unless **LAPT-PC** is regarded as useful for dissemination, the links to this system should be left off (see recommendation 6.2.2).*
- *Authoring system user and technical support details such as Problems/Errors/Suggestions etc. should be buried under links -- this could also be reconstructed as FAQs.*

#### Recommendation 6.2.2

*Make a clear separation between the LAPT website which appears to be mainly designed for local student support (and is also closely associated with Professor Gardner Medwin's everyday teaching and research website) and one which is designed for dissemination, has a teacher focus and supports remote authors.*

#### Suggestion 6.2.1

*Give some consideration to a name for the system which is more indicative of what it is than LAPT. Several interviewees found this meaningless. This will need some reflection as LAPT clearly has a considerable following beyond the institutions which use it regularly. However a name change/relaunch might be good excuse to revisit possible supporters with news of updates etc.*

### 6.3 Explanations of LAPT and demonstration questions

#### Recommendation 6.3.1

*Rewrite the bullet points which should:*

- *Make a clear distinction between student focus and teacher focus (different pages would be good).*
- *Make a clear distinction between " why use LAPT?" and "how to use LAPT." taking them from one to the next.*
- *Use only assertions which are clearly evident -- see insert.*

- *Make assertions which address learning needs. For example:*

*"If you use this system you will:*

- *get a clearer idea of your strengths and weaknesses*
- *be able to target your revision more accurately*
- *develop the skill of knowing how confident you are*
- *be better equipped for decision-making in the real world*
- *etc*

*... something you would not have been able to do with a standard marking system."*

**For instance the authors found:**

*"To encourage you to think about how reliable your answer is." -- is convincing whereas*

*"To encourage you to think about how your answer relates to a wider range of things you know about." -- is not immediately demonstrable.*

*The statement "To penalise you and wake you up if you are making confident wrong answers -- to make you think!" -- is clearly correct whereas the assertion "To improve your study habits -- question where your ideas come from and how reliable they are." -- it's not immediately clear that LAPT will do this.*

**Recommendation 6.3.2**

*Avoid the use of graphs in this top-level succinct explanation -- the general feedback seems to be that they are not helpful except to a minority of statistically numerate individuals.*

**Recommendation 6.3.3**

*Give examples of the application of confidence based marking which relate to real life. Although it seems obvious that an accurate knowledge of confidence or the lack of it is a universal life skill, it might be a good idea (at first) to use examples where this is most critical.*

*For example:*

*"It is vitally important that when a doctor makes a decision which involves the health or life of a patient, he or she is aware of how confident they are. For instance, if you are a junior doctor on the ward late at night and a decision needs to be made about necessary treatment for a patient, if you are confident about the decision, then the action taken will be to treat, whereas if you know that you are not confident in this particular knowledge/decision, the action to be taken will be to wake the consultant up.*

*This system, apart from being a better way of assessing your knowledge and skills, will have the secondary effect of training you to be aware of your level of confidence at all times."*

*Then follow this up with 10 to 20 well-written questions which illustrate this assertion.*

**Recommendation 6.3.4**

*Use a limited number of demonstration questions on the site which have the following characteristics:*

- *They are clear and well written.*
- *They demonstrate the assertion that a knowledge of confidence is important.*
- *At least some demonstrate that MCQ's can be written to test deeper and more complex learning and are not just useful for testing superficial knowledge.*

**Recommendation 6.3.5**

*Provide an attractive menu driven set of authoring support pages around the question authoring tool which clarify the do's and don'ts of using the tool, the manual and provide access to FAQs and troubleshooting pages.*

**6.4 Pedagogic research and support**

**Recommendation 6.4.1**

*It will be important to highlight the fact that, although CBM is a system of negative marking, it also allows students to avoid negative marking when their confidence is low. This makes LAPT a fairer system. Making this clear will go far to combat the concerns in some circles with negative marking.*

**Recommendation 6.4.2**

*With respect to the dissatisfaction with multiple true/false questions in medical circles, it might be a good idea to give some thought to:*

- *concentrating on more widely accepted question types such as Extended Matching Questions (EMQ) and Single Best Answer Questions*
- *discussing/devising a marking system which was able to reward and punish appropriately without being perceived as negative marking.*

### **Recommendation 6.4.3**

Several interviewees felt unsure about using the system for summative testing. Some of this centred around the setting of pass marks and some around student desire to practice tests in the same format as the summative ones they can expect. Some clear advice on the setting of pass marks, supported by literature would help potential adopters.

#### **Suggestion 6.4.1**

It might be a good idea to do some research to find out whether or not there is any evidence base for the assertions that:

- objective questioning only tests superficial knowledge (there is apparently a considerable body of literature to support ordinary objective questioning)
- negative marking is a bad thing
- confidence based marking introduces gender bias (the authors think that Professor Gardner Medwin has clearly proved the contrary but it will be useful to be clear where this assertion comes from).

#### **Sources for further information on pedagogic aspects**

One of the interviewees suggested that the Higher Education Academy subject centre for Medicine, Dentistry and Veterinary Medicine (<http://www.itsn-01.ac.uk/>) might be a useful first start as they may have been looking into issues in objective testing.

Another interviewee suggested the National Foundation for Educational Research (<http://www.nfer.ac.uk/>) as being a possible source of useful information -- in a later e-mail communication he attached a number of other useful references and contacts. This extra information has been communicated to Professor Gardner Medwin.

#### **Suggestion 6.4.2**

Devise, if possible, some research to demonstrate that CBM actually helps the learner when contrasted with conventional marking. It is appreciated that there may be potential ethical problems in devising control groups.

#### **Suggestion 6.4.3**

If a prominent educationalist could be persuaded that this version of CBM is a good idea, arguing the case will be considerably easier.

## **6.5 Adoption of standards**

### **Recommendation 6.5.1**

Work towards conforming to the IMS-QTI standards and make it clear that a system for the importation of compatible questions from other formats exists. This should allow bulk importation of questions from other systems (the authors are aware that Professor Gardner Medwin currently provides a question importation service). This suggestion is tentative as the authors themselves are not technically familiar with these standards – the suggestion comes from one of the interviewees.

## **6.6 Commercial or academic partnership**

### **Suggestion 6.6.1**

Further investigate the development of associations/partnerships with commercial CBA products which, if they included CBM as an option would ensure its dissemination over a wider area (the authors are aware of the association with Speedwell Computing).

The UK-based Course Genie (<http://www.coursegenie.com/>) has been suggested -- an advantage of this is that the producers of Course Genie have a good reputation for adopting user suggestions, a disadvantage might be that currently, Course Genie does not have a numeric scoring system at all (although in conversation recently, there was a suggestion that a scoring system is about to be released).

### **Suggestion 6.6.2**

Investigate/reopen the development of a relationship with the Open University. Although this institution has its own internal inertia, it is comfortable with the idea of objective testing and, more importantly, one of the interviewees is enthusiastic about LAPT. In conversation, he seemed to imply that the introduction of confidence based marking by the OU had initially failed because of poor timing and that investing in this relationship further might produce results. It goes without saying that the adoption of this system by the OU would be extremely significant.

## 6.7 Institutional inertia

### Suggestion 6.7.1

*One possible driver is the recognition by the Quality Assurance Agency inspectors of a need to constantly re-examine and upgrade assessment systems -- they look for this. Timing in respect of the next QAA visit could be important.*

### Suggestion 6.7.2

*Study the socio--politics of institutional change. A tongue in cheek suggestion! This is outside the remit of LAPT dissemination and indeed of this evaluation. It would however be very interesting to look at the socio-politics of institutional change in HE to see how the dissemination of a potentially good academic idea/system is best achieved. Might be a suggestion for a PhD.*

## 6.8 Areas to target or avoid

### Recommendation 6.8.1

*There are some areas where the adoption of LAPT might be easier or more difficult. Areas to target might be those such as medicine and engineering where a knowledge of confidence is self-evidently important. Alternatively, there are areas such as high-stakes summative testing where there is a political necessity to demonstrate the "safety" of LAPT in terms of validity over and above what has already been done.*

## 7. Further evaluation

The above findings represent the opinions and reactions of 13 key individuals all of whom are involved in teaching in higher education. To formally confirm them, it would be necessary to perform more extensive studies with larger numbers of staff and students as well as to initiate further conversations with educationalists.

In spite of this, the authors feel that there is sufficient agreement between themselves and the interviewees to make the findings of this study helpful in the further development and dissemination of CBM/LAPT.

## 8. Conclusions

The authors have been impressed by the extensive use of CBM within a limited number of institutions. Personal contact, seminars and demonstrations seem to work well -- as does support for authoring by the developer. The concept of CBM is readily understood even in verbal communication and everyone interviewed would like to see it adopted more widely.

The web site currently works satisfactorily as a support system for students and staff already using the system but is less good at encouraging new adopters at remote sites. Although improvements to the website will be useful, it may be that its principal function will be to back up the primary "sales effort" which comes from demonstration and personal contact. Nevertheless the website is an important secondary element in that someone who is persuaded to try CBM and sell it to colleagues/ administrators, needs a very clear easy-to-use website which gives succinct and persuasive explanations backed up by well-written examples.

The authors were interested in what appears to be a degree of confusion around the whole field of objective questions and negative marking and feel that some of the problems in disseminating the CBM system owe more to fashion than to evidence -- if this is the case, it might well be possible both to present convincing arguments and to avoid particular problem question types such as Multiple True/False.

Personal and institutional inertia will remain a problem although a policy of supporting local champions might help with the initial establishment of CBM in a remote institution. Developing a version of CBM which works within one of the popular commercial systems might also get round this. Finally, the authors would recommend a further exploration of the relationship with the Open University as being potentially rewarding.

It is interesting to note that, such is the power of the basic idea, this review itself has proven to be a useful instrument of dissemination, stimulating potential new users and encouraging previous enthusiasts to revisit the concept.

The complex of academic research, innovation, ideas, software tools, and procedures which is the London Agreed Teaching Protocol represents an enormous investment of ideas, time and creativity on the part of its developer, Professor Gardner Medwin. For it to move into the mainstream it now needs to come under the "ownership" of more than one individual. This might come about through the creation of partnerships with organisations such as the Open University or through partnerships with "champions" in other institutions or by adoption into one of the commercial authoring packages -- or maybe a combination of these. Any of these routes should provide the impetus and investment necessary for the next stage of its dissemination.

## 9. References

**Gardner Medwin & Gahan 2003** [\*Formative and Summative Confidence-Based Assessment\*](#) A.R. Gardner-Medwin & M. Gahan (2003) Proc. 7th International Computer-Aided Assessment Conference, Loughborough, UK, July 2003, pp. 147-155

## Appendix 1 -- Interview Technique

Potential interviewees were divided into three categories\*:

- individuals who had never heard of LAPT -- termed **Naive**
- individuals who had examined LAPT but had not generally adopted it -- termed **Non-Adopters**
- individuals who regularly use LAPT -- termed **Adopters**

*\*a later interview took place with a clinical educator*

Interviewees (and the authors) were teachers in higher education, in all cases having some experience of e-learning. Disciplines included pathology, microbiology, physiology, mathematics, engineering mathematics, learning technology and public health.

Interviewees were initially contacted either personally or by e-mail.

The Naive group were told about the concept of Confidence Based Marking and asked to look at the LAPT website for as long as they were motivated. They were interviewed, either face-to-face or by telephone on their reactions to the concept, the system and the website.

Adopters and Non-Adopters were also interviewed, by telephone or face-to-face, in more detail about what they found attractive and helpful about the system, including the website, what they found unhelpful and their reasons for adopting it/not adopting it. They were also asked how they had come across the system and what had motivated them to engage with it.

The clinical educationalist was asked specific questions about the pros and cons of objective question types and negative marking.

## Appendix 2 -- Sample question file used by LAPT -- appearance to author and user.

(Copied from the instruction manual)

Here is a complete but simple file that would work fine. It has just one SECTION, with several fairly obvious question types (described in more detail below). Note that the functional parts of the file can be adorned with comments. Note that lines beginning // and everything following // on a line are ignored by the program.

```
S("Questions about Britain"); // This heading will be shown with all the questions in the section.

// True/False Question
Q("Britain is in the Northern Hemisphere",T); // The specified answer is True.

//Text answer Q
Q("What is the capital of England?"); // A question requiring a text answer
A("London"); //This is one acceptable answer, and is shown as the model answer
A("Londres","Yes, but the English name is London"); //a different acceptable answer with a special
explanation
A("lond* "); //accepts anything beginning with lond (NB by default, the checking of the reply is
case insensitive).

// A multiple choice Question (the students chooses just one of the options as correct)
Q(MCQ,"Which is the capital of Scotland?");
M("Glasgow");
M("Edinburgh");
M("Stirling");
A(2); // the correct answer is the second option listed above
I(1,"This is not the capital, but it is the largest city"); // this picks up a particular incorrect response
(option 1) and gives a special explanation when it is chosen.

// A multiple response Question
// NB the student is asked to select two options in this case, one at a time, and with confidence
expressed for each choice.
Q(MRQ,2,"Which two of these towns has a cathedral?");
M("Chester");
M("Preston");
M("Gloucester");
M("Banbury");
M("Bath");
A(1); A(3);

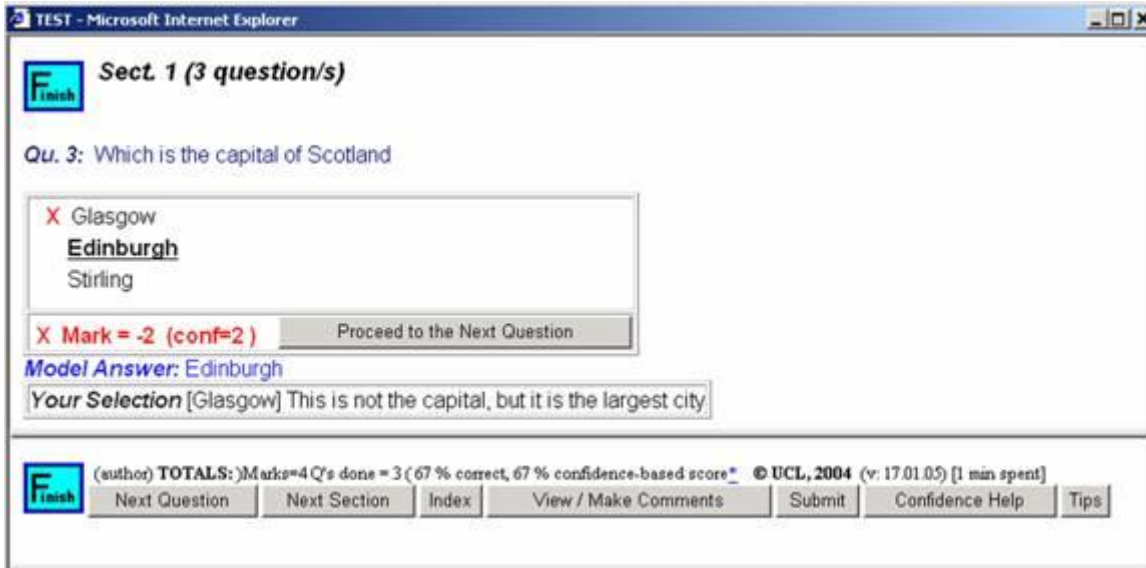
// A Multiple Response Q marked all in one go. Confidence is expressed that the entire set of choices
is correct.
// NB the Q text could include, if you want, an indication of how many ticks the correct selection would
include
Q(MRQ,0,"Make an exact selection showing which of these towns is on the coast");
M("Birmingham");
M("Southampton");
M("Leeds");
M("Liverpool");
M("Nottingham");
A(2); A(4);

// Another text question, but with answers that may be given in numeric form.
Q("How many separate countries or provinces make up the UK?");
A("4"); // model answer
A("four"); //allows text version
A("{4} * "); // A bit unnecessary in this case! - but would allow as correct '4.0 countries' (any format
of 4 (±1%) with or without a following word).
E("They are England, Scotland, Wales and Northern Ireland: The 'United Kingdom of Great Britain
and Northern Ireland' . It does not include the Channel Islands or the Isle of Man but for VAT
purposes the Isle of Man is treated as part of the UK. "); // Explanation shown after the student's
response has been marked.
```

//A question with a range of quantities (number + units) permissible as correct answers  
Q("British shops sometimes still use the 'pound' as a unit of mass. How many kg is one pound.?");  
A("0.454 kg"); // The model answer  
A(" {0.45 0.46} \* "); //gives a permissible range, and the wild card \* allows any or no unit, since the unit was specified in the question.  
A(" {450 460} g\* "); //also accepts if a student gives the correct answer, but in grammes

Most LAPT files are actually simpler than this one - typically they have a lot of questions, and perhaps many sections, but with only a few question types.

**The screen grab below** illustrates the appearance to the student of the question highlighted above -- the student having selected an incorrect answer. As can be seen, it includes feedback appropriate to the answer. The student has answered incorrectly at confidence level 2 and has thus had two marks deducted.



## Appendix 3 -- Discussion of gender bias in multiple-choice questions (from the website of LTSN-01)

[I have heard that multiple-choice questions \(MCQs\) are biased in favour of males. What is the evidence for this?](#)

[http://www.ltsn-01.ac.uk/resources/faq/index\\_html?my\\_num=23#23](http://www.ltsn-01.ac.uk/resources/faq/index_html?my_num=23#23)

The largest controlled study looking at the topic of MCQs and gender bias was conducted by the Educational Testing Service (Cole, 1997) over four years with 400 different tests and more than 1500 data sets involving millions of students ranging in age from Grade 4 (9 years old) to graduate school, including those taking the MCAT (Medical College Admissions Test). A clear finding was that asking students to *produce* the answer rather than *select* the answer (constructed response versus multiple choice format) did not produce different gender effects when the *same* question was asked in *different* formats. (Other studies with the same results include De Mars, 1998; Hamilton and Snow, 1998; Ryan and Fan, 1996.) However, in the constructed answer questions, answers requiring written responses favoured females and ones requiring the production of a figure or the interpretation of graphical information favoured males. This kind of content effect may be of more importance than the question format.

Other studies also indicate that it is primarily the content and not the format that shows a gender effect. For instance, a study (Ryan and Fan) of 6000 14 year old students on an international mathematics test showed that males scored better on algebra and geometry, but females performed better on arithmetic, though all questions were presented in MCQ format. A similar result from the ETS study found females performing better on computation and males on conceptual questions (though the differences were extremely small in real terms).

A study (Hamilton and Snow) of tests in biology and astronomy taken by 1100 17 year olds showed no format effect when comparing constructed format and MCQs, but males scored better on questions involving spatial or visual content. Interestingly, a short lesson on spatial-mechanical reasoning eliminated the effect. This could indicate a difference related to experience rather than an inherent gender traits. This is also suggested by a study by Byrnes et al. (1997) in which maths items that showed a large gender difference in American students showed no such difference in Chinese students.

Some studies have found gender differences with multiple-choice questions favouring males, the most thorough of which is Breland (1991). However, this study looked at the Advanced Placement Tests in History in the US, a test that is taken by students in advanced courses in the subject who are, in general, high ability students. This presents another possible complicating factor.

A consistent finding in the ETS study is that scores in representative samples of the general population invariably show larger spread for males than females - more high scorers and more low scorers. If stable differences are indeed seen in vet, medical or dental students, this could be a factor, as these populations are likely to be self-selecting for higher ability students. If the range of scores is greater for males in general, then any high ability sample is likely to magnify a male performance advantage because of the higher proportion of high scorers relative to the female sample. This does not indicate a biased test necessarily, but reflects a restricted sample from the statistical distribution of scores in the general population. (See Cole, 1994, for a more thorough discussion.)

**Conclusions:** Current evidence at most suggest a weak systematic gender bias for MCQs, though there are potential complicating factors. For instance, there is some indication that high ability students, which includes most medical, dental and veterinary students, show greater gender effects with males scoring slightly higher on MCQs and females on constructed response items. There is also evidence that 'well-known' gender differences, e.g. in maths and science, are significantly smaller than 30 years ago (Cole). Again, changes in educational experiences may be decreasing previously observed differences. Also, it is critical that questions, in any format, are well-written, reliable and actually measure what is intended. (See NBME handbook by Case and Swanson for very useful guidelines on writing many types of questions.)

**Caveat:** Many of the studies cited here were conducted on high school students in the USA. The oldest students were of approximately the same age as first year students in UK universities - 17-18 years old - and so the results are at least somewhat likely to be applicable. So far, I have found very few (0) published studies of gender differences in test items involving UK medical, dental or vet students. A more complete review of this topic is underway. Please let me know ([jean@ltsn-01.ac.uk](mailto:jean@ltsn-01.ac.uk)) of studies that are not included here, particularly of medical, dental or vet students, ideally in the UK.

## References for Appendix 2:

- Breland, H. M. (1991). *A Study of Gender and Performance on Advanced Placement History Examinations*. College Board Report No. 91-4. 44p.
- Byrnes, J.P., Hong, L. and Xing, S. (1997). Gender differences on the math subtest of the Scholastic Aptitude Test may be culture-specific, *Educational Studies in Mathematics*, v34, n1, pp. 49-66.
- Case, S. M. and Swanson, D. B. (2001). Constructing written test questions for the basic and clinical sciences: Third Edition. National Board of Medical Examiners. Available at: <http://www.nbme.org/nbme/itemwriting.htm>
- Cole, N. (1997). *The ETS Gender Study: How Females and Males Perform in Educational Settings*. ETS Technical Report, Available at: <ftp://ets1.ets.org/pub/res/gender.pdf>
- DeMars, C.E. (1998). Gender differences in mathematics and science on a high school proficiency exam: the role of response format, *Applied Measurement in Education*, 11(3), pp. 279-99.
- Hamilton, L. S. and Snow, R. E. (1998). Exploring differential item functioning on science achievement tests, CRESST Report 483.
- Ryan, K.E. and Fan, M. (1996). Examining gender DIF on a multiple-choice test of mathematics: a confirmatory approach. *Educational Measurement: Issues and Practices*, 15(4), p. 15-20.
- Author: This FAQ was written by Dr Jean McKendree and does not reflect an official endorsement by the LTSN or any other organisation. Any questions or comments can be sent to: [jean@ltsn-01.ac.uk](mailto:jean@ltsn-01.ac.uk)