Work in Progress:
LaTeX Open Format Testing (LOFT) System for Multiple-Choice Quizzes and Exams

George Stetten

University of Pittsburgh, Carnegie Mellon University, stetten@pitt.edu

Abstract – **The LaTeX Open Format Testing (LOFT) system enables the composition and automatic grading of paper-based quizzes and exams. A standard format for multiple choice questions permits individual questions and answers to be composed in the widely-used open-standard LaTeX typesetting environment, thus permitting a very broad range of notation and inclusion of figures. Automatic grading using any standard inexpensive scanner with a multi-sheet feeder is supported. Answers may be permuted for each student to discourage cheating. The system removes any dependence on centralized facilities for procuring and scanning score-sheets or storing test questions. The goal is to facilitate collection and maintenance of a body of multiple-choice questions for a given course or area of the curriculum, which can be presented for study to all students, supplemented by the continual addition of new questions. The format for question generation is extremely simple and stable, avoiding the risk of depending on proprietary software, which may change over the long term invalidating the significant work of compiling questions. Other features include the ability to add errata to questions after discussion with students and to accept alternate answers as correct. Results are emailed automatically to individual students immediately after scanning.**

Index Terms – automated grading, bubble sheet, multiple-choice exams, LaTeX.

TEST DESIGNS

Multiple-choice exams have long been the mainstay of standardized tests such as the SATs and are widely used in many educational settings. They are inherently fair, in that they are graded objectively, and save considerable time in grading especially with large classes. Their downside includes a lack of individual instruction that is possible only with hand grading and restrictions on the sorts of questions that can be asked. The objective of this article is not to debate these issues, but rather to present a particular method of implementing multiple-choice exams with potential advantages over others – namely control over the entire process from test and answer sheet generation to automatic scanning of answer sheets, and a format that is open, versatile, and inherently stable over the long term.

The LaTeX Open Format Testing (LOFT) system has been under development for the past three years, during which time it has been in continual and successful use in two large required undergraduate courses in the Department of Bioengineering at the University of Pittsburgh. Specialized software has been developed to manage the creation and grading of multiple-choice tests. Two particular formats have been used, (1) a full-period exam with up to 38 questions printed in a multipage format with a separate "bubble sheet" answer form, and (2) a single sheet quizz with three questions in which the bubble sheet is on the same page as the questions. These both are produced using a standard laser printer. Permutation of the question order is supported so that students cannot simply copy from each other's answer sheets. Typically, three versions of the full-period exam are created and then copied, stapled automaticallly, and collated before handing out. The one page quizzes are printed directly on the laser printer, each one being a different permutation from each of the others and identified by an extra set of circles filled in by the printer. Class lists are supported, each student being given a short ID number to use on the answer sheet. Email addresses are stored for reporting grades, allowing the daily quizzes to provide immediate feedback and encourage attendance.

This paper briefly describes the LOFT system and the custom software we have developed in the Java programming language, which we call "JavaLOFT". An example of a three-question quiz from our course in electronic circuits is included to show how tests are created, graded, and the results shared with the students.

SOURCE CODE FILES

The heart of the LOFT system is a standard format for the definition of a test and for each of the questions in the test. Two types of "source" files are used to do this, each using only standard ASCII (American Standard Code for Information Interchange) characters to guarantee accessibility without any proprietary software. All personal computers come with some kind of simple text editor that is capable of editing ASCII files, as these files contain no hidden formatting characters. Programming in languages such as C++, Java, or HTML typically is performed using such ASCII files to define the operation of programs. HTML, in particular, is known as a "mark-up" language,

FIGURE 1
TOP PORTION OF 3-QUESTION QUIZ GENERATED FROM THE MCT FILE (FIG. 2), MCQ FILE (FIG. 3), AND A CIRCUIT
SCHEMATIC (CIRCUITS0004.JPG) LINKED TO FROM THE MCQ FILE (FIG 3).

because it basically formats how a page will appear when "run" on a certain system (in that case, a "browser"). Another such "language" is LaTeX, a widely-used open source typesetting standard in which postscript and *pdf* files (printer description files) are fully specified and created from ASCII source files. LaTex is known for its versatility and stability, and is favored by researchers who do not wish to be at the mercy of commercial word-processing and formatting software [1]. The formats for the two types of source files we have developed rely on LaTeX to define how questions will apear on the test.

To demonstrate the usage of LOFT, we present the example shown in Figure 1, which shows the top portion of a 3-question quiz on electronic circuit theory including the first question and the answer section with bubbbles filled in by the student, who has entered his/her identification number and answered all three questions (including two questions not shown). Bubbles identifying the particular permutation order of the questions are also shown, pre-filled by the printer (black circles), as well as fiducial markers (3 diagonal pairs of squares) used by the image processing software to automatically score the exams.

The first type of source file in the LOFT system (see Figure 2) always has a filename with the suffix '.mcq' (multiple choice question), and includes snippets of LaTeX code defining a particular question. This includes the question itself, the possible answers (with A being correct and B-E being incorrect), and an explanation to be used for the answer sheet (see Figure 5).

Each section in a '.mcq' file begins with a identifier line starting with `%_MC` (LaTeX comments always begin with the `%` character and as such are ignored by LaTex). Thus, in Figure 2, the line

    %_MC question

identifies the section of LaTeX code defining the question and

    %_MC answer A

identifies the section of LaTeX code defining answer A, the correct answer. Note that the answers are defined in an unpermuted order, and are only permuted later by the software when the tests are produced.

An explanation section of LaTeX code headed by

%_MC explanation

is included for use in the production of an answer sheet for the exam, from which students can learn by reviewing their own tests or studying using questisons from previous years. Two other sections (not shown) may be included for "errata" and "alternates", to explain why credit has been given for one of the other answers.

```
%_MC question
What is the resistance between points A and B,
if $R_{1}$ = 10 M$\Omega$, $R_{2}$ = 40 $\Omega$, and
$R_{3}$ = 40 M$\Omega$? Calculate only to two significant
digits.
\vspace{.2cm}

\includegraphics[width = 2 in]{circuits0004.jpg}

%_MC answer_A
10 M$\Omega$

%_MC answer_B
30 $\Omega$

%_MC answer_C
50 M$\Omega$

%_MC answer_D
30 M$\Omega$

%_MC answer_E
40 $\Omega$

%_MC explanation
The total resistance is $R_{1} + \frac{R_{2}R_{3}}{R_{2}+R_{3}}$.
However, this can be simplified. Because $R_{2} << R_{3}$,
$R_{3}$ can be ignored, the total resistance becomes $R_{1}
+ R_{2}$.  But it can be simplified even more.
Since $R_{2} << R_{1}$, $R_{2}$ can be ignored,
and the total resistance is simply $R_{1}$.
```

FIGURE 2
QUESTION (MCQ) FILE DEFINING A PARTICULAR QUESTION, THE CORRECT (A) AND INCORRECT ANSWERS (B-D), AND AN EXPLANATION.

The second type of source file in the LOFT system (see Figure 3) always has a filename with the suffix '.mct' (multiple choice *test*). This file defines which questions will be imported (which .mcq files) in what order on the test, as well as the test title, date, and format (currently full-period exam or 3-question quiz). It also includes a

%_MC permutations

section, which states how many tests will be printed (65 in our example) and what permutation "key" to use (1327 in our example). Permutations are generated using a standard pseudorandom number generator, which is initialized using this key to ensure a unique order of permutations for a given test. With a maximum of 5 answers permitted, a given question can have $5\times 4\times 3\times 2\times 1=120$ different orders for the answers. Thus an exam of n questions can have 120^n possible permutations, essentially infinite. We limit the actual number to 3999.

OPERATION

The 3-question quiz is typically handed out at the end of class, often containing material that was covered in that very class. This encourages students to pay attention and ask questions when they do not understand something. After class, the quizzes are immediately fed through a scanner in the instructor's office, and the scoring program (described below) automatically reads the bubble sheets and emails the results to the individual students.

```
%_MC import
circuits0113.mcq
circuits0112.mcq
circuits0111.mcq

%_MC permutations
65
1327

%_MC bubble3

%_MC title
BioE 1310 - Quiz 1

%_MC date
1/7/2010
```

FIGURE 3
TEST (MCT) FILE, SPECIFYING WHICH QUESTIONS ARE INCLUDED IN A TEST, THE NUMBER OF TESTS AND PERMUATION KEY, TITLE, DATE, AND TEST FORMAT (THIS IS A 3-QUESTION QUIZ WITH BUBBLE SHEET INCLUDED ON A SINGLE SHEET).

Figure 4 shows a typical email that would be generated from our example. In it, Joe Student (ID = 24) can surmize that he got 2 out of 3 questions correct on Quiz 1 on January 1, 2010, and that the mean of the 54 students who took the quiz was 1.9 with a standard deviation of 0.4. He can look on the class website to see the answer sheet with an explanation (see Figure 5), whose answers are unpermuted He can see in his email that the particular answers he gave were A for question 1, b for question 2, and E for question 3. Capital letters are used for correct answers (A and E) and lower case for incorrect answers (b). Normally, all correct answers are A, but if an ambigous question is identified (often by a particularly attentive student), the instructor may automatically give credit for alternative answers. In this case, answer E in question 3 is receiving credit.

```
From:      stetten@pitt.edu
Subject:   BioE 1310 - Quiz 1   1/7/2010
Date:      January 1, 2010  02:15:39 PM EST
To:        joestudent@pitt.edu


ID = 24  score = 2/3   AbE   mean=1.9   SD=0.4  N=54
```
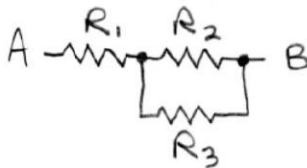
FIGURE 4
EMAIL GENERATED TO STUDENT FROM SAMPLE EXAM

BioE 1310 - Quiz 1   1/7/2010
Answer Sheet - Correct answer is A for all questions

1. What is the resistance between points A and B, if $R_1 = 10$ MΩ, $R_2 = 40$ Ω, and $R_3 = 40$ MΩ? Calculate only to two significant digits.

A. 10 MΩ

B. 30 Ω

C. 50 MΩ

D. 30 MΩ

E. 40 Ω

Explanation: The total resistance is $R_1 + \frac{R_2 R_3}{R_2 + R_3}$. However, this can be simplified. Because $R_2 << R_3$, R_3 can be ignored, the total resistance becomes $R_1 + R_2$. But it can be simplified even more. Since $R_2 << R_1$, R_2 can be ignored, and the total resistance is simply R_1.

[*circuits0113.mcq*]

FIGURE 5
ANSWER SHEET GENERATED FROM

The scoring program also produces a spreadsheet, readable by Microsoft Excel or other standard spreadsheet programs, containing the scores of all the students as well as basic statistics such as mean, standard deviation, etc. These can be easily imported by the instructor into a master grade sheet for the entire course.

BUBBLE SHEET READER

The standard form used for automatic grading, the "bubble sheet", is nothing new, although free software for grading them is not readily available. The most prominent company in the field is Scantron, Inc. [2], whose scanners are expensive enough to usually require they be shared centrally in a school. At our university, at least, this means that every time a test is to be graded, the sheets must be carried to the centralized facility, where there is a delay of perhaps a day before the grades may be picked up. For daily quizzes, in particular, this adds up to an appreciable extra time load. Another limitation of standard answer sheet scanners is that they only use standard answer sheets, and therefore cannot accommodate custom single-sheet quizzes in which the answer form is included on the same page as questions.

Solutions with more flexibility are available commercially. Optical Character Recognition (OCR) software, which comes with the purchase of most scanners, is generally limited to interpreting the printed word, but specialized software may be bought for producing and scanning custom multiple choice answer sheets (e.g. FormIdea, from Tech Vision, Inc. [3]). To our knowledge, however, a reliable freeware solution for creating and scanning customized answer sheets is not yet available, especially one integrated into on open format system for generating tests with the full power of typesetting offered by LaTeX.

Our automated answer sheet scoring system begins with scanning the answer sheets using a standard multi-sheet scanner (HP LaserJet 3055, from Hewlett Packard, Mountain View, CA, ~$500-600) creating 75 dpi color (RGB) image files in the jpeg format. These are stored in a directory where the JavaLOFT software automatically scores them, interpreting the pencil marks using conventional image analysis techniques. Variations in page location, orientation, and scale are accomodated using fiducial marks on the printed form to apply an affine transform to the image, and mean pixel intensity within the circles is interpreted using

and adaptive threshold. The system has not exhibited any errors from answers properly entered on the answer form ("keep your pencil marks within the circles") in three years of use in two large undergraduate courses (~50 students each), although we are basing this on student feedback, assuming we would be told if the tests were not graded accurately. Two exams have completely checked by hand, yielding a perfect score for each in terms of the program's reliability. The only ambiguity noted thus far has been in cases where a student does not fill in any circles for a given question, in which case the automated system occasionally interprets this case as having one of the circles significantly darker than the others, instead of no answer being given. Erased answers have been correctly graded.

DISCUSSION

A number of competing technologies exist to paper-based multiple choice exams. Particularly popular are the "clickers", portable devices held by each student that communicate via infrared or radio to a central computer and give students the individual ability to repsond to questions, typically displayed using a projector. A popular manufacturer of such equipment is Hyper-Interactive Teaching Technology [4]. At present, clickers have the significant drawback compared to paper exams that they do not permit each student to go at his/her own speed through all the questions on the exam. However, with the availability of ever-less expensive "netbooks" (small notebook computers) and tablet computers, in the near future such equipment could be handed out to individual students, affording the advantages of instantaneous grading eliminating paper altogether. The underlying test structure and archives developed using paper-based systems today could still, of course, be of value.

The LOFT system provides significant time savings in grading, not just compared to conventional exams, but also compared to multiple-choice exams using the standard Scantron answer sheets. These needed to be transported to a centralized facility on campus where a 24 hour turnaround time was typical. With the LOFT system, grades are typically emailed to the students from the instructors office within the hour. Combined with immediate posting of the answer sheets, this encourages feedback from students about ambiguous answers, resulting in automatic credit to any student who choses that same answer. The quality of the archive of questions is also improved in this way. In our classes, we post all old questions for review, to eliminate any advantage students may have by saving exams from one year to the next. Thus the archive receives ever further scrutiny, with students often identifying ambiguous questions while studying for the exam.

The time it takes to prepare questions with the LOFT system is essentially the same as it would be using a conventional word processor, at least when the questions consist solely of words; LOFT is perhaps a even bit faster, because formatting the questions and answers is not required. Inserting figures is very simple, also comparable to a standard word processor. Equations take a bit longer if the operator is not familiar with LaTeX. Devotees of LaTeX who use it regularly claim that creating equations is faster than with a graphical equation editor such as that embedded in Microsoft Word®. But for those who do not use LaTeX regularly, there is a significant learning curve, and finding a particular symbol and debugging may take a frustratingly long time. There are a number of free online graphical equation editors that produce LaTeX, and these can help. On the flip side, a key advantage of LaTeX is its comprehensive assortment of mathematical symbols, if one knows how to find them. Google® has made this job easier.

ONGOING WORK

To address the learning curve of LaTeX, we are planning to create an interface that will convert questions created with Microsoft Word® to LaTeX. Planned efforts also include improving the user interface and developing a website-based version for use via a browser. We further envision a system for the capture of hand-written answers in a special section of the answer sheet, which can be stored for efficient presentation to a human grader, or conceivably even graded automatically. We expect to introduce more sophisticated statistical analyses of the test results, and develop complete documentation to encourage use by others than ourselves.

Since the LOFT system represents an open format for tests generation, anyone can write software to manipulate the .mcq and .mct files to create and grade multiple choice tests. Our efforts using the Java programming language will be posted on-line at **www.javaloft.org**. It is our hope that this may lead to the development of archival libraries for individual courses, across multiple courses in a particular curriculum, and even between institutions in similar fields of study.

REFERENCES

[1] http://www.latex-project.org/

[2] http://www.scantron.com/, Irvine, California.

[3] Tech Vision, http://www.tkvision.com/product/idea.htm

[4] http://www.h-itt.com/index.html

AUTHOR INFORMATION

George Stetten, M.D., Ph.D. Professor of Bioengineering, University of Pittsburgh; Research Professor, Robotics Institute, Carnegie Mellon University; stetten@pitt.edu, www.vialab.org; www.javaloft.org.

George Stetten earned an A.B. in Engineering and Applied Physics from Harvard in 1976, an M.S. in Biology from NYU in 1986, an M.D. from the SUNY Syracuse in 1991, and a Ph.D. in Biomedical Engineering from UNC, Chapel Hill in 1999. Dr. Stetten wrote the software for the first

computer system on board Deep Submersible Alvin at the Woods Hole Oceanographic Institute, and he designed a telemetric egg to study incubation of endangered birds at the Bronx Zoo that was nominated for the 1992 Discover Magazine Award For Technological Innovation. He conducted the first full-scale classroom in which laptop computers were networked wirelessly in 1994, helped

develop the first Real-Time 3D ultrasound scanner at Duke University, was a founding contributor to the National Library of Medicine Insight Toolkit (ITK) for image analysis, and invented the "sonic flashlight," a device for image-guided medical procedures.