Honours Project Report

ScriptView

Victor Soudien

Supervised By: Gary Stewart

Mark Table

# Introduction

The management of test scripts at tertiary institutions involves collection and distribution to markers, evaluation and redistribution to students. The current methods in place rely on the physical test scripts, require a considerable amount of time and coordination between multiple individuals. These methods increase the waiting time for students who ideally need to get feedback as soon as possible in order to improve for future assessments. Current test management also lacks the collection of any data about the tests aside from the student’s final mark. This leads to a lack of historic data for educators to reference when creating new tests.

Needs more.

# Background

This section presents the relevant findings from previous and current work that either directly or indirectly influenced the design and development of the system.

## Functionally similar software

Despite the increasing prevalence of technology in education, the marking of paper-based tests has remained largely unchanged. There has however been significant work done on trying to digitize the marking process which shows some promising results.

One of the most widely recognized techniques is that of Optical Mark Recognition (OMR) which utilizes optical equipment to interpret marks on a paper [1]. These marks are usually indicated in a grid format with cells corresponding to letters or digits. Commercial products which implement OMR are available and have had limited success. One of the key constraints of OMR is that it limits the examiner in terms of what type of questions they can ask. This is because OMR is only suited to multiple choice which implies that examiners can only ask questions up to the applying level in Bloom’s taxonomy.

An alternative to this would be to use a system that could analyze free form answers. This kind of technique has been explored, especially in terms of essay marking.

Researchers such as Christie [2] have found success applying automatic marking to the marking of essays. The researcher describes a technique used to mark both style and content which requires acceptable style metrics to be set up before marking as well as manually marking a certain amount of scripts to calibrate the algorithm. Such a system would allow for questions from all levels of Bloom’s taxonomy and would therefore be more widely applicable.

Pulman [3] did similar work in that he attempted to use machine learning techniques to automatically mark short free form answers. Although he draws no conclusions on how easy his method is to customize to different questions, there is a problem which is identifiable in both his and Christie’s approaches. That problem, is training data. A system that is more generally applicable to a wide range of test is preferred since time does not have to be spent on customizing it for each new type of test.

There are also certain issues which could arise which Thomas [4] expanded upon in his work. Thomas performed an experiment in which he compared the results of a fully automated marking system and a manual marking system. The tests intended for automatic marking were taken on a computer and the other tests were done on paper. The researcher noted that certain problems could arise during the taking of the test that could only be solved effectively by human markers.

The most problematic of these being that of ambiguity, either in the student’s answer or the examiner’s question. Students may also not be able to express themselves clearly, especially if they are taking the test in a language other than their first language. Both of these problems could be overcome by manual markers who used their knowledge of the domain and discussion with fellow markers to assign appropriate marks.

## Forms of assessment

An advantage of electronically processing tests is that it creates the opportunity for the collection of large amounts of data about students’ progress. This data can then be analyzed to predict learning patterns and activities that might indicate a risk of failure [5]. As one of the main goals of the system is to allow for the recording of such data, it was essential to understand what kind of information an educator may look for and how it would be used.

It was found that tests are often divided into two distinct categories namely formative and summative assessment. Formative assessment is used during the course of a particular section such that the results can be used to inform how the student or educator should progress [6]. Summative assessment on the other hand is used at the end of a section to gauge the extent of the student’s learning for grading, certification or the evaluation of the effectiveness of the teaching method [6].

The terms formative and summative assessment however, do not describe the format of the tests but rather their function. As a result, a range of question types can be used in both. A classification of these types based on their cognitive complexity was developed in 1956 by B. Bloom and revised (see figure 1) in 2001 by L. Anderson [7]. The cognitive level of complexity increases as one ascends from remembering to creating. The taxonomy provides a framework which educators can use to ensure that they are testing all aspects of a student’s understanding as well as allowing them to create tests which assess the most appropriate levels for the given situation.



Figure 1: Original and Revised Bloom's Taxonomy [2]

Aboulsoud [8] claims that formative assessment is more valuable to students since it provides them with feedback which they can use to improve their future work thus increasing their chances of academic success. He suggests that for formative assessment to have this effect, it should be followed by immediate feedback which clearly indicates areas in which the students should improve while acknowledging their effort to achieve the objectives of the assessment.

This implies that any system developed to aid the marking process should allow for rapid feedback with the characteristics outlined by Adoulsoud as well as allow for the marking of questions at various levels of Bloom’s taxonomy.

## Converting physical tests to digital format

The biggest bottle neck in the entire back-end process appears to be the scanning of test scripts so that they can be processed and stored electronically. This was noted by Doctor H. Suleman (personal communication, 24 April 2014) of The University of Cape Town who has implemented a system which involves the scanning of marked test scripts to automate the process of capturing marks on the university’s learning management system known as Vula. The system he developed only requires that the cover page of the test be processed since it is the one which contains the marks and student information. However, all pages are scanned since they will be emailed to the student as feedback. Due to this, the quality of the scanned document needed to be high enough so that the student would still be able to read the feedback written by the marker and that the image processing algorithms could successfully detect the student number on the cover page which was indicated by shading certain pre-defined areas.

To accomplish this the documents were scanned at 300dpi but it was decided that the scanning was too slow and thus proved infeasible. To improve the speed of scanning, the resolution was changed to 200dpi and the documents were scanned in black and white. This sufficiently reduced the scanning time while maintaining the readability of the document.

A problem that was not addressed by Dr. Suleman during our interview, was the size of the scanned documents. This was however addressed by Doctor J. Tangkuampien (personal communication, 1 May 2014), who also scanned documents at 300dpi and had to reduce it to 200dpi.He however, reduced the resolution due to concerns about the file size. When scanning at 300dpi the file size was approximately 4MB and decreasing the resolution halved this size. Another difference between his scanning solution and that used by Dr. Suleman is that he scanned the documents in colour. While scanning the documents in black and white would have further reduced the file size, Dr. Tangkuampien noticed that doing so interfered with the visibility of the text if students used highlighters. This is because when scanning in black and white, very light highlighters such as yellow do not show and dark ones obscure the text.

Dr. Suleman used a combination of OMR and manual name entry to name scanned files whereas Dr. Tangkuampien used a fully manual approach. With the first approach a special cover page needed to be created whereas with the second the existing test format could be used.

S. Chetty (personal communication, 5 May 2014) described a method which uses optical character recognition (OCR) to name files. This method required that a certain area of the script be reserved for the student number, which eliminated the need for a dedicated cover page as the area for the student number could be included on the same page as the first question of the test. This is essential as he suggests that a large part of the process will be a compromise between the existing process and the new electronic solution. By eliminating the creation of a cover page the overall process can also remain as simple as possible.

Minimizing file size while maintaining the readability of the final document is essential to the final system as limited storage is available; scripts for multiple tests need to be stores and image processing needs to be performed on the documents.

## Software interaction

In order to make the best use of the tablet and stylus interface, it was necessary to understand what users have found intuitive in the past.

Alisi [9] discusses the concept of natural interaction during the description of systems implemented to improve the experience of museum visitors. The researcher suggests that natural interaction can “reduce the gap between computing and ordinary physical things” however it would require that the interfaces differ from traditional human-computer interaction such as the use of menus and icons.

The Point At system described allows the user to point at a character in a picture in which they are interested in and the system will provide them with more information on that character. Since this a natural action and similar to how the user would traditionally ask a tour guide for information, it allows the technology to become a transparent medium instead of overwhelming the experience.

This point is further expanded upon by Malizia [10] who believes that a natural interface, especially gesture based ones, should allow users to interact with the software using the same gestures they would use with the actual physical object. The researcher states that users should not have to learn an arbitrary set of gestures to use the software but instead the software should allow gestures which take the user’s habits, background and cultural aspects into consideration. The researcher recognizes that this goal might not be achievable but knowledge of these aspects means that we can design interfaces that are as natural as possible by carefully considering what will be intuitive for the target audience.

In order to provide the students with useful feedback, Dr. Tangkuampien suggests that a three-pane view is the most useful. The three-pane view describes the screen and document layout used that presents the viewer with the question, student answer and model answer in a single view. During his research Dr. Tangkuampien has found that this view has proven the most beneficial to both the marker and the student. It allows the students to easily assess what they did incorrectly and how they can fix it. The software shown during the interview maintained the traditional view of a test script i.e. as a set of multiple pages exactly as they were scanned with the model answer added to the right column

Dr. Suleman observed that when working with electronic test scripts one does not have to maintain this view. He suggests that it is possible and may be beneficial to tailor the view of the test script to the device on which it will be marked. When changing the view however, the ability to add feedback should still be maintained as Fowles [11] suggests that this is both reassuring to the marker but is also essential if the marking is challenged in the future.

# What is ScriptView?

ScriptView is an application which aims to improve the current test management procedures by leveraging the advantages of technology at specific steps in the process. This allows the application to be more familiar to users and thus improve their acceptance of it.

# References

* 1. Poor, D. D. (1995). Image Capture and Storage Techniques in Association with Optical Mark Reading
  2. Christie, J. R. (1999). Automated essay marking-for both style and content. *Proceedings of the Third Annual Computer Assisted Assessment Conference, Loughborough University, Loughborough, UK*
  3. Pulman, S. G., & Sukkarieh, J. Z. (2005). Automatic short answer marking. *Proceedings of the Second Workshop on Building Educational Applications using NLP,* 9-16.
  4. Thomas, P. (2003). The evaluation of electronic marking of examinations. *ACM SIGCSE Bulletin,* *35*(3), 50-54.
  5. Siemens, G., & Gasevic, D. (2012). Guest editorial-learning and knowledge analytics. *Educational Technology & Society,* *15*(3), 1-2.
  6. Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal,* *22*(5), 537-548.
  7. Forehand, M. (2010). Bloom’s taxonomy. Emerging Perspectives on Learning, Teaching, and Technology, , 41-47.
  8. Aboulsoud, S. H. (2011). Formative versus summative assessment. *Education for Health (Abingdon, England),* *24*(2), 651. doi:651 [pii]
  9. Alisi, T. M., Del Bimbo, A., & Valli, A. (2005). Natural interfaces to enhance visitors' experiences. *Multimedia, IEEE,* *12*(3), 80-85.
  10. Malizia, A., & Bellucci, A. (2012). The artificiality of natural user interfaces. *Communications of the ACM, 55*(3), 36-38.
  11. Fowles, D. (2005). Literature review on effects on assessment of e-marking.