Honours Project Report

ScriptView

Victor Soudien

Supervised By: Gary Stewart

Mark Table

# Introduction

The management of test scripts at tertiary institutions involves collection, distribution to markers, evaluation and redistribution to students. The current methods in place rely on the physical test scripts and 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 on 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.

Advancements in this fields often focus on the creation of a completely automated electronic solution. While these have found limited success, their greatest weakness is in their inability to handle ambiguity either in the student’s answer or the educator’s questions. Other solutions such as those which utilize optic mark recognition require institutions to change their current procedures to accommodate this method by forcing educators to use a predefined format for the test or limiting the types of questions they can ask.

Through the development of an electronic test management system which combines elements of technology and human marking, this research aims to answer the following questions…

Needs more. Research questions.

## Project Overview

ScriptView is an electronic test management system which aims to enhance the current test management procedures by leveraging the advantages of technology at specific stages in the process. These advantages include the rapid processing of large amounts of data, automated error checking and the ability to efficiently maintain a digital paper trail. It is a complete test management solution which aims to reduce the time taken to mark and redistribute tests as well as capture and analyze the results.

The system allows users to scan the paper versions of the tests and mark them on either a web or tablet interface. The scanned tests are automatically sorted and stored on a central server according to the course and test name. This server will be access controlled in order to ensure the security and integrity of the tests.

Once the tests are available on the server, they can be marked on either a web or tablet interface. Each of these interfaces are optimized to offer the most intuitive marking experience given their limitations such as screen size and input type. Both interfaces aim to decrease the time that is required to mark a test while maintaining high level of accuracy.

ScriptView allows for the distribution of marked scripts to the students and for a summary of the marks of a particular test to be sent to a teaching assistant (TA). The format of the email sent to students allows them to quickly assess in which areas they lost marks and thus where they need to improve. The summary sent to the TA is in CSV (Comma-Separated Values) format which is compatible with most learning management systems and thus reduces the time needed to capture these marks on another system.

## Report Format

This report details the design, implementation and evaluation of the mobile marking application as well as the components of ScriptView which involve image processing such as, the processing of the scanned scripts and the memorandum.

# 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.

## The current test management procedure

The information presented here was gathered through observation of the current procedures. These observations were conducted at the University of Cape Town (UCT) during the researcher’s time as an academic tutor.

Students write the paper based tests which are then collected by the TA. The TA distributes these tests to the tutors which are responsible for marking them. The tutors have to gather at a single location at a given time to mark the scripts so that any ambiguities can be discussed and clarified. The tests are divided equally among the tutors and each has the responsibility of marking the entire test and writing the total marks awarded on the first page of the test.

Once all the tests have been marked, they are returned to the TA who captures them onto the university’s learning management system. The marked tests are placed at a collection point within the department building and the students are notified via email that they can collect their scripts.

An important observation was that students often neglect to collect their scripts or the scripts are misplaced. This entire process from the point of collecting the test scripts until they are available to the students can take up to 5 days. However, it should be noted that this is because the UCT Computer Science department puts an emphasis on quickly returning scripts to students. In general the turnaround time is much longer.

## 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.

## Summary

Summary of background work

Diagram for system and what I’m doing

# Design and Implementation

## Software Development Methodology

A software development methodology is a framework which guides the planning, implementation and evaluation of a software system [12]. The two most well-known frameworks are the waterfall and agile methodologies. While there is variation within these two categories, the central theme of each variation remains the same.

The waterfall methodology places an emphasis on planning and performing all tasks in a sequential manner as can be seen in Figure 2. The idea is that by performing thorough planning at the beginning of the project, changes can be avoided at later stages. There is also a strong emphasis on having only a single implementation phase in which the entire project is developed. The advantages of this approach are that the planning and resulting documentation can often be beneficial to inexperienced teams and it allows for the strict monitoring of the progress of the project.

Where the waterfall approach aims to mitigate having to incorporate change at later stages of the project through planning, the agile methodology embraces change. This is done by going through multiple iterations of the phases shown in Figure 2 excluding the maintenance phase. The advantages of this approach are that it allows developers to respond to changing user requirements - as long as these remain within the scope of the system; it promotes the involvement of users within the process thus ensuring that the final product better meets their expectations and it allows for the creation of user testable products throughout the development process.

Since ScriptView is intended to enhance a current process, the feedback of users involved in the current operations is essential. This meant that it would be necessary to create versions of the final product which could be used for user testing throughout development. It was thus decided that following the agile development methodology would be the most beneficial.

Figure 2: An overview of the phases of the waterfall methodology

A software design methodology is … (Mention waterfall and Agile; also touch on OOP, high cohesion and low coupling)

## Programming Paradigm

Throughout development an Object Oriented (OO) approach to system implementation will be followed. The OO programming paradigm states that components should be thought of as objects with an emphasis on their properties, behavior and interaction with other objects [13]. This fits the ScriptView system well as each of the various components may be written in different languages but they will need to communicate in order to function correctly.

Even though this communication will be necessary, all objects created will aim to have low coupling and high cohesion. Coupling describes how dependent one object is on another and cohesion describes how related elements within an object are [14]. By having low coupling and high cohesion within the system, it will be more reliable and maintainable [15].

## Evaluation Design

### Data Collection

Include questionnaires in appendix and mention here. User centric design stuff should go here

The effectiveness of the system will be evaluated by using seven tutors from the CSC1010H Computer Science course. Each tutor will be required to mark three test scripts. One will be marked as they normally would, without the system, while the other two will be split between the two interfaces.

The marking will be taking place in a controlled environment in the sense that outside interference will be limited to only the course convener, TA or examiner. The tutors will be marking both sets of test in the same location and it will be a location which is familiar to them.

Upon completing both sets of tests the participants will be asked to complete a survey (See Appendix A). Based on the feedback changes will be made to the system before another test is conducted under these same conditions. After which another survey is completed.

The results from these two surveys is what will be analyzed in order to answer the research questions.

### Data Analysis

Most responses are given using a Likert-scale so it would be appropriate to use mode or median [16] to draw conclusions about the data. The mode and the second most frequently occurring value will be used to determine how divided the responses were. This is because if the mode is 4 and the second most frequent is 1, then it means that tutors either strongly agreed or strongly disagreed and therefore the reason behind this should be further investigated.

## Process Design

ScriptView is intended to enhance the existing test management procedures as described in section 2.1. This means that the system needed to be designed in such a way that it would not disrupt the current operations to a large degree.

Figure 2 shows the overall process flow of ScriptView. This flow was inspired by the existing test management systems but decreases the amount of human intervention required.



Figure 3: Overall process flow of the ScriptView system

The first step in the process is preprocessing. The preprocessing step ensures that all constraints on the physical test paper (see section X) are adhered to before the scripts are scanned. These checks are performed by the user but it is not time consuming since the appropriate templates for the test papers are provided and if these are used, this step can be skipped by the user.

Once scripts have been collected and scanned, they are automatically emailed to a monitored inbox. As scripts arrive at the inbox, they will undergo Optical Character Recognition (OCR) in order to determine the name of the course and the test so that they can be stored on the server. This was done to remove the need for users to manually enter this information for each test which decreases the amount of work by the user as well as reducing the probability of naming errors since all text output from the OCR methods is checked a database of test and course information. In addition to the OCR, the digital copies of the tests are also undergo image processing in order to improve their readability when they are displayed on the tablet interface. Once both the OCR and image processing has been completed the script is uploaded as a set of images, each representing a single page from the script.

Ethical concern: Image processing on tests

The act of marking a test script is largely unchanged in terms of where it fits within the process flow and how it is performed. This was done intentionally in order to decrease the amount of changes that would have to be made to the existing test management procedures if an institution wishes to adopt the use of ScriptView.

Test marks are automatically gathered, stored and analyzed by ScriptView. These marks can then be accessed by administrative users and sent to students along with their test scripts. Marks can also be exported in Comma-Separated Values (CSV) format for uploading to the institutions learning management system. (How is this different from the ScriptView section?)

## Initial Software Design

Since an agile methodology with a focus on user feedback was followed during the development of ScriptView. The final product is, in some respects, vastly different from the original design. Therefore this section only documents the initial design of the software before any implementation or user feedback sessions was undertaken.

The components of ScriptView discussed in this paper can be clearly divided into three sections namely test script processing, memorandum processing and the mobile marking application.

### Test Script Processing

Test script processing involves all tasks that are required scan the scripts; retrieve information from the scripts for storage; converting them to a suitable format for the marking interfaces and converting them into a suitable format for emailing to the students.

Two approaches were considered for scanning the test scripts. The first was the creation of a dedicated desktop application which would allow the user to select a set of tests from the local storage on the machine and upload these to the server. This would allow the operation to be strictly access controlled and thus easily traceable to a single person. The disadvantages of this approach however is that the scripts would first need to be scanned and saved to the local machine, which is not only time consuming but creates a single point of failure. The uploading would also be delayed if the user forgets to start the processing on the local machine and becomes preoccupied with other administrative tasks.

The second approach was to have the scripts automatically emailed to a monitored inbox as they were scanned. This allows for the work to be distributed among multiple individuals by allowing different people to do the scanning while still keeping the processing centralized. The process would still be traceable as each documents has a timestamp that can be checked against who was using the scanner at that particular time. The machine doing the processing, is still a single point of failure but the scanned scripts will not be lost as they will still be available in the email inbox and will be processed as soon as the applications resumes. This approach was chosen as the final design for the scanning of test scripts.

The file structure on the server required that the course and test name of a scanned script be known. Since scripts are automatically emailed from the scanner, this information cannot be retrieved from the user at the point of scanning. It was decided that the best solution to extracting this information, would be to perform OCR on the scanned test pages.

Once the OCR has been completed the test needs to be uploaded to the server. It is at this point that it has to be converted to a format which is suitable for the marking interface. It was decided that the tests would be uploaded to the server as a set of images in Portable Network Graphics (PNG) format. This would allow for easy access to a single page within the document but was also influenced by the capabilities of the target device. The decision to do the conversion at this point was influenced by the limited computational power available on the target device for the mobile marking application (See section 4.5.3). It also improves the speed at which the mobile application can provide access to the test scripts to the user.

Before tests can be emailed to the students, the sets of images need to be converted to individual Portable Document Format (PDF) documents. This was designed as a standalone application which will be available to the students.

The diagram in Figure 3 shows the flow of a single test script through the system. It was used to ensure that the script was always in the expected format at a given stage.



Figure 4: The flow of a script through the system

The diagram in Figure 3 was used to create the analysis class diagram in Figure 4. The aim was to design classes that would be as decoupled and cohesive as possible so that if changes needed to be made to any of the existing stages in Figure 3 or a new one added, then it would have a low impact on the number of classes that would be affected.



Figure 5: Analysis class diagram for script processing component

The EmailMonitor class handles all tasks which involve the email inbox this includes downloading new emails as they arrive as well as downloading the PDF attachment to be send to the PDFProcessor. The PDFProcessor will split the PDF into a series of images as well as perform the OCR on the document to determine the course and test name which will be used in the upload directory. The FileUploader will construct the upload directory, which is the test script’s destination on the server, and upload the images.

### Memorandum Processing

In order to create the mobile marking application, the questions and answers of the test is required. To avoid having the user enter this data, the memorandum will be uploaded to the server via a web interface. It will be processed by a standalone application which will need to extract the questions, answers and answer regions. The answer regions are the start and end y-coordinates of the space left open for the student to answer. For more detail on why this is required see section X.

To accomplish this the memorandum will be split into multiple pages, similar to what is done to the scanned script, and OCR will be performed on each page. String comparison and text position will be used to categorize extracted text into question and answer categories. The position of the answer text will then be used to extract the answer regions.

### Mobile Marking Application

The mobile marking application is the core of the three ScriptView components discussed. It is the only part that is exposed to the user and thus its design needed to be focused around usability and increasing marking speed.

As mentioned in Section 4.6.3 the target device for the mobile marking application runs on the Android operating system thus the application was designed with the Android design guidelines [17] in mind. The principles which have the greatest influence on the design are discussed below:

1. **Real objects are more fun than buttons and menus**

This principle states that it is better to allow users to directly interact with objects in the application than to have them use buttons or menus. This principle directly influenced the decision to allow the user mark the test script through the use of gestures. Using the stylus, the user is able to annotate the test script using one of the gestures shown in Figure 5 below. Each of these already has a meaning to the marker as these are the annotations they would have made had they been marking the physical test paper. From left to right the annotations indicate a single marks, half a mark and no marks.

Insert image of tick, half tick and X

1. **Keep it brief**

This principle states that instructions to the user should be kept as short as possible. Since the users of the system will have had some experience with marking test scripts. The instructions need only inform them about how to accomplish what they would usually do on paper and not the technicalities of marking. There will also not be a lengthy tutorial on how to use the application but rather tooltips for graphical elements. This guideline also influences the design of error message as the aim is to keep them brief and user friendly.

1. **Only show what I need when I need it**

Tasks which the user cannot perform in a given context will be hidden while those that are possible will be shown using an icon rather than being hidden in a menu. This guideline also applies to the use of tooltips as the user will only need to see this additional information when they need it.

1. **I should always know where I am**

This says that users should be confident about where they are within the app and how to navigate to previous screen. This has inspired the use of tabs at the top of the application for each of the questions that need to be marked. This will allow the user to easily move between questions beyond just moving to the previous or next question.

1. **If it looks the same, it should act the same**

This principle ties into the concept of recognition versus recall. Which states that it’s faster to interact with an object if you recognize its purpose than if you have to think about it. To aid recognition standard Android icons will be used for common actions like undo and redo. Since the user encounters these in multiple other applications they will easily recognize their function within the marking application. However, care must be taken in ensuring that the correct actions are tied to the correct icons as an incorrect mapping will cause frustration and may lead to users becoming reluctant to use the application.

While always keeping these concepts in mind the design of the application started at analyzing the various tasks a user may want to perform when using the application. To facilitate this a use case diagram was used as shown in Figure 5.



Figure 6: Use case diagram of mobile marking application

These use cases where then grouped according to how often they would be used together to ensure that all functionality on a given screen was related and would adhere to the third android design guideline stated above. Once use cases had been grouped, the first screen that was designed was the one that users would use to mark a script. This was prioritized since it is the screen that the user uses for the longest time and thus having it ready for user testing would be beneficial.

Based on the findings from the interview with Doctor J. Tangkuampien the marking screen was designed to offer the marker a view of the question, answer and student test script (Hereafter referred to as the three-pane view). This three-pane view (see Figure 6) would increase the markers marking speed as it would always have the current question and answer in view thus decreasing the amount of work required by the marker. A mockup of the design of the marking screen can be found in Appendix B.1. As can be seen in the diagram, tabs were used to allow the user to navigate between questions and the largest area of the screen was dedicated to the student answer since it is the focus of this screen.

****

Figure 7: Three-Pane view which served as a guide for interface design

The next screen which needed to be designed was the script selection screen. The purpose of this screen is to allow the user to select which test script they wanted to mark. The challenge presented by this screen was that it needed to show mainly navigational information such as the test and course name. Appendix B.2 shows a mockup of the design of this screen. The decision to use an Android Navigation Drawer layout was influenced by the third Android design guideline since it enables the application to hide the course information when the user is looking at tests. This was the same motivation behind using an expandable list view for the test information as scripts for another test could be hidden or shown at the users command.

The final screen that was designed was the script upload screen, a mockup of which can be found in Appendix B.3. The purpose of this screen is to allow the user to check their mark allocations, name the script according to the student number and upload the script. As can be seen in the diagram a table was used to display the mark summary as this is a familiar format to the marker and thus improves usability. The screen also displays a scaled down version of the test to allow the user to verify that marks on the test paper were accurately counted.

All screens were designed to make use of the Android Action Bar, an example of which can be seen in Figure 7. This is a bar along the top of the screen which contains the application name, action items and an overflow menu. The decision to use the action bar with familiar icons for each action item was influenced by the fifth Android design principle.



Figure 8: An example of an Android Action bar that includes (1) app icon, (2) two action items, and (3) action overflow [18]

The state diagram in Figure 8 shows which screen a user needs to be presented with at a given time. This diagram was used throughout development to ensure that data for a given screen was available at the correct times and that a user could only reach a certain screen from specific locations.



Figure 9: State diagram used to indicate screen flow of the ScriptView mobile application

## Constraints

ScriptView is intended to enhance an existing procedure and as a result certain constraints needed to be established. This was to ensure that the system would be easily adopted by potential users as well as decrease the amount of changes that would have to be made to their current workflow. This section describes theses constraints and how they arose.

### Scanned Test Script

During the design of ScriptView, sample test scripts were provided by the project supervisor. These scripts were previous tests that were used in a single course. Figure 5 shows the variation that exists in the formatting of the first page of each of the scripts.



Figure 10: Variation on cover pages

This variation meant that the scanning component of ScriptView would need to be able to handle all of them as well as any unseen cases. To accomplish this OCR was performed on each cover page and the resulting text was searched for keywords. These keywords were related to the information which needed to be extract for example name, surname or test name. This approach had limited success. The location of fields that needed to be completed by the test taker such as name and surname were accurately detected but the handwriting recognition which was then performed on these regions was unreliable and inaccurate. Text that was typed by the educator such as test name was accurately detected by the OCR but determining whether or not text was part of the test name required that some intelligence be incorporated into the algorithm and this was beyond the scope of the system.

In order to overcome these challenges it was decided that the formatting of the first page of the test would need to be constrained. This would allow the OCR algorithms to categories recognized text based on their position on the page and relative to each other. An example of this is that the algorithm could now determine that text after the institution name is the name of the course.

A disadvantage of this approach was that the student number was no longer detected. Since the student number is essential to the system a compromise had to be made. This compromise was that the system requires the user to enter the student number before uploading the test. This compromise was acceptable as it promoted blind marking (define and reference blind marking) and checks could be coded to ensure that the student number was accurate.

Keeping the goal of reducing changes to the current procedures in mind one of the sample tests was chosen as the accepted format and it only constrains the top half of the first page. The chosen format can be seen in Figure 5 highlighted in green.

### Memorandum

The memorandum for any test that will be marked using ScriptView needs to be uploaded via a web interface in PDF format. This enables the system to manipulate the way in which the questions and answers are displayed on the marking interface. In order to accomplish this, the text of both the questions and answers needed to be extracted as well as the location of the answer regions.

Extracting the text was accomplished using a similar technique as was used on the first page of the test. However, instead of using OCR the text is simply extracted from the document. This is possible since the memorandums must always be typed and are not scanned documents. Since the length of both questions and answers could vary and neither the question nor answer positions were known categorizing the text as on the first was not effective.

To mitigate this problem, the use of annotations on the memorandum was explored. This approach involves having the educator place annotations at certain location on the test paper which would indicate the start and end of an answer region. The annotation that was used to indicate the start of an answer was the mark indication which was already a part of the test. To indicate the end of answer a new annotation would need to be created. An example of these annotations can be seen in Figure 6.



Figure 11: A question and answer with the appropriate annotations

These annotations allowed for algorithms which could accurately categorize text into either being a question or answer based on their location relative to the annotations. These annotations were also used to extract the start and end y-coordinates of the answer regions which were needed for mark allocation on the marking interface.

Although this technique was effective, through testing, it became apparent that the spacing on the memorandum and the test script that is handed to students is not always identical. This meant that the coordinates for the answer regions which were extracted would not always be accurate and would as a result skew any analytics performed on the test results.

To avoid this inaccuracy it was decided that the user would need to upload both the memorandum and the test script as it is handed to the students, both in PDF format. Using these it was possible to categorize questions and answers without the use of the annotations as answers were not present on the test handed to students so categorizing text could be achieved by comparing the extracted text from both documents. Answer region detection is now performed on the test as it is handed to students by using image processing to detect the lines left open for a student to write the answer.

The use of annotations was thus discarded but it was now necessary for the user to upload two documents instead of only the memorandum and the space left for students to answer needed to be indicated by lines as seen in Figure 7. Figure 7 also indicates the accepted location for the mark indication. In an effort to reduce the amount storage required, only the extracted text and region coordinates are stored and not the uploaded files.



Figure 12: The accepted format for a question

### Target Device

The final version of the mobile marking application needed to run efficiently on a Samsung Galaxy Note 10.1 and use the Samsung S Pen stylus as input. The operating system on the device was Android Version X (Jelly Bean).

The key constraints introduced by this device was making effective use of the 10.1” (255.8mm) screen and managing the memory when handling a large amount of images.

### Summary

Is this necessary?

## Project Planning

With four months to fully implement the system and the goal of involving users in the development process, a clear plan needed to be established for all phases of the project. Since the agile methodology was going to be used, the project was divided into four iterations. At the end of each of the first three iterations, there would be an opportunity for potential users to provide feedback on the current state of the system (See section 4.3). This feedback could be on any aspect of the system irrespective of whether or not it had been fully implemented.

Tasks within each iteration were prioritized based on their importance in terms of achieving the overall goals of the software such as improved marking efficiency and reducing the probability for error. User feedback also influenced the priority of tasks as highly requested, yet in scope, features were prioritized.

## Project Feasibility (First Iteration)

The first iteration of developments began with an assessment of project feasibility. It was decided that to demonstrate that the system could be completed in the given time, the most complex components would need to be prototyped. These components included the image processing which needed to be performed on the scanned test scripts and the recognition of gestures made using the Samsung S Pen.

During the prototyping of the image processing algorithms the capabilities of the Open Source Computer Vision (OpenCV) library was tested. OpenCV is an image processing library with a strong focus on efficient, real-time applications and provides the user with a range of optimized image processing algorithms such as the Canny Edge Detector. Development of the prototype was done in the C++ programming language as it allowed for the fastest execution times and would thus be suitable for the task of processing tests as they were scanned.

In order to assess the feasibility of the OCR that needed to be performed on the regions detected through the image processing, Tesseract was used. Tesseract is an open-source OCR engine [19] which is compatible with a range of image formats. At this stage of the project the Linux command line tools available through Tesseract were used and no code using the library was implemented. This was efficient to prove the feasibility of using the Tesseract engine on the scanned documents.

The gesture recognition component was coded using Java with the Samsung S Pen Software Development Kit (SDK). The SDK allows the user to make full use of the S Pen within any Android application. By using the SDK it was possible to create and save gestures to represent a tick, half tick and cross. The focus then shifted to allowing the user to perform multiple gestures on a page and have all of them detected individually as the SDK only allowed for the detection of a single gesture on a given screen.

There was no user testing at the end of this iteration but the prototypes were presented to the project supervisors before continuing to ensure that the requirements were understood.

Scanning at 300dpi

## Core Development

Development on what would become the final product began after the prototypes from the feasibility iteration were accepted by the project supervisors. The section presents the three iterations which followed as well as how the user testing at the end of the first two impacted the design of the system. Only the results of the user testing relevant to the implementation are discussed here, a more detailed analysis of the results can be found in Section X.

### Second Iteration

The second iteration began on the 21st of July 2014 and ended on the 10th of September 2014. The aim of this iteration was to complete most of the core functionality that would be required to perform user testing and get constructive feedback. Thus the components that were prioritized during this iteration was the scanned test script processing, the memorandum processing and the core marking interface.

The first task that was performed during this iteration was to move all reusable code from the feasibility iteration to a repository on GitHub. GitHub is a code-hosting repository based on the Git version control system [20]. From this point onwards it was used to manage all code produced during production of the system in order to allow for effective change and bug tracking.

Work began on the script processing component by implementing techniques which would allow for the monitoring of an email inbox. To accomplish this the JavaMail Application Programming Interface (API) was used. The API provides a framework to build mail and messaging applications and was used to establish a secure connection to a mailbox and download any emails that were received including any attachments. In order to prevent malicious use of this application, it can only be executed when it is on the institutional network and is connecting to a mailbox which is part of the institution. An example of this is that the UCT Computer Science issued email addresses ([example@cs.uct.ac.za](mailto:example@cs.uct.ac.za)) can only be monitored using the application if it is running on a computer connected to the UCT network.

In order for the test script to be uploaded to the server, it had to be converted to a series of images, where each image represents a page. Since the OCR to determine the course and test name would require the first page to be converted to an image before being processed, it was decided that converting all pages to images before performing the OCR would decrease the time taken to process a single script. This was especially apparent with larger scripts as it only required a single iteration through all the pages of the script. To convert the pages to images, Apache PDFBox was used. It is a Java API which allows for the manipulation of PDF documents and was thus ideally suited for the task as the scanned scripts emailed as PDF Documents.

The next step in the script processing was to extract the relevant information from the cover pages. As mentioned before, this is information is the course and test name. In order for this component to work seamlessly with the email monitoring component, it was decided that it should be implemented in Java as opposed to C++ which was used during the feasibility iteration. Since work had already been done on investigating the capabilities of the Tesseract OCR engine, Tess4j was used. Tess4j is an API which allows access to the Tesseract API to Java applications. The increase in processing time when moving from C++ to Java was investigated and was found to be negligible due to the limited amount of work required by the Tesseract API.

The final element of the script processing component was to upload all the images and memorandum text to the server. The design class diagram in Figure 11 shows a detailed view of the class structure of the script processing component.

Insert design class diagram of script processing

The majority of time in this iteration was focused on development of the mobile marking application. Development of the mobile application began at the marking screen. This is the screen that users would use most often and it was expected that it would be the one to change the most based on user feedback. Completing it in time for the user testing at the end of this iteration was therefore essential.

One of the biggest challenges faced during the implementation of the mobile application was implementing the three-pane view discussed earlier. Since the design mockup in Appendix B.1 was done to scale, the challenge of effectively using the limited screen space had already been solved. The implementation challenge however was arranging the various Android layout components, with their own limitations on size and position, to achieve the desired effect.

The following features where also implemented on the main marking screen, the ability to undo an annotation, the ability to leave a comment and page navigation using tabs at the top of the screen. To allow the user to navigate the memorandum buttons were added to the bottom left of the screen. The appearance of the marking screen at the end of this iteration can be seen in Figure 12.

Insert screen at iteration 2

The both the screen to select a test to mark as well as the screen to upload a marked test were implemented. One of the challenges of the test selection screen was populating the lists presented to the user with the appropriate test available on the server. At this point in development the course and test names where retrieved from server files structure and not the database. This was accomplished through the use of the Java Secure Channel (JSch) API. The API enabled the java application to securely connect to and browse the server using secure shell (SSH).

The JSch API was also used to enable the user to name the test on the final screen by renaming the folder on the server. The API also facilitated the uploading of the marked images to the server. The appearance of the test selection and script upload screen at the end of this iteration can be seen in Figure 13. Once the script had been uploaded methods needed to be created to release the memory used for images of the current test script.

Insert test selection and script upload screen at iteration 1

At the end of this iteration a user testing session was conducted. Fifty percent of the users stated that the test script was unreadable, this was in reference to the image quality and not the students’ handwriting. Fifty seven percent of users also rated the effectiveness of the screen layout three or below which means they were undecided or though it was ineffective. This meant that the three-pane view discussed in section 3.5.3 was not proving to be as effective as initial thought. This is emphasized by the fact that eighty five percent of users felt that the interface did not improve their marking speed.

In terms of interaction, seventy one percent of user said that it was easy to mark using the tablet interface which is supported by the fact that eighty five percent stated that marking with the stylus was intuitive. Many users did ask what was meant by intuitive and it was stated to them as how natural it was to use the stylus, this is did they have to give it much thought or not. This meant that no significant changes needed to be made to the stylus input other than tweaking parameters to make the gesture recognition more accurate.

### Second Design Iteration

Based on the feedback received at the end of the previous iteration with regards to script readability and screen layout, it was decided that a second design iteration would be undertaken for both as the two are not mutually exclusive. This is because the size dedicated to displaying the script directly influences its readability and any processing which can be performed on the script before uploading to the server decreases the size it needs to be displayed at on the mobile application.

The second design iteration for the script processing resulted in the addition of an image dilation stage which aims to thicken the lines of both the typed and written text thus improving its readability. Since performing image processing on the scanned scripts raises some ethical concerns around manipulating the students answer, the original unprocessed version of the students’ script is also uploaded to the server. The flow diagram in Figure 14 shows the updated flow of a script through the system.



Figure 13: Updated script flow diagram

The three-pane view was reevaluated and the comments from users was used to guide the new design. The most obvious improvement was to increase the size dedicated to displaying the test script and upon further investigation it was discovered that it was not necessary to display the question at all. This is because users are accustomed to reading the question on the student’s script and then looking for the answer in another location. By placing the question on the screen as well, the application was actually negatively impacting the marker by showing them information they were not interested in. This realization led to the discarding of the three-pane view and a change to the template seen in Figure 15.

Users noted that they did not find having to move between memorandum answers using the next and previous buttons intuitive and would prefer if the application showed only answers which were relevant at a given time. To accommodate this, the buttons at the bottom left of the screen were discarded and the memorandum display was changed to show only the answers for a given page. The display was also designed such that users could select a given answer and the test would automatically scroll to that section of the student’s test script. As can be seen in Figure 15 the memorandum answers where now designed to have clear divisions between each answer since all answers for the current page would be displayed at the same time.



Figure 14: Marking View template after first user feedback

### Third Iteration

Once the second design iteration had been completed the implementation of these changes was prioritized to ensure that they would be complete in time for the final user testing session. In addition to these changes, the mobile application was also connected to the database in order to retrieve course and student information. Connecting to the database was achieved through using PHP Hypertext Processor (PHP) code which was hosted on the server. Using PHP allowed for highly reusable code which could be accessed at various points in the application.

Connecting the application to the database enabled the development of predictive text and rigorous error checking when the user names the test scripts which greatly reduces the possibility of incorrectly naming a script. With the application connected to the database functionality such as uploading the student’s mark, flagging a script for follow up and access control could be implemented.

In order to collect information on mark allocation which would be beneficial to the analytics which needed to be performed, it was necessary to determine to which sub-question a mark was assigned. This was accomplished by taking the median y value of the annotation and determining if it falls within a certain answer region. If the annotation was not assigned to a sub-question then it was matched to the region to which it was closest. By using this method an accuracy of 95% for mark allocation was achieved. A visualization of this method can be seen in Figure 15.

Insert diagram of mark allocation to sub questions

A problem introduced by this method however was that it slowed down the gesture recognition. Even though this increase in processing time was not noticeable to the user, it did have implications for the accuracy of the final mark. This is because gesture recognition is not performed on the main application thread and multiple gestures could be recognized in parallel. This meant that one thread could read and update and old mark value which would then be overwritten by another thread which had also read the old value, essentially discarding one of the updates. To solve this problem, access to the marks was restricted to a single method which used the Java synchronized keyword. This prevented multiple threads from accessing the mark value at the same time by causing threads to block if another thread was modifying the mark value.

With changes to the mobile application complete, improvements to the script processing needed to be made which only became apparent after using it during the first user testing. These improvements included having to handle scripts on which the OCR was not successful or scripts that were incorrectly uploaded to the server. Both of these errors were mitigated by using a backlog of scanned scripts, an upload window and the database of course information. The upload window was defined as the time during which a successful upload had occurred. The state diagram in Figure 16 shows how these two components were used.

Figure 15: State diagram showing use of backlog and upload window

A second user testing session was conducted at the end of this iteration. Since the application was more feature complete for this session, there were more User Interface (UI) elements on the screen and as a result user feedback was focused around interaction and feedback. Many users stated that their crosses were being counted as ticks. Users also expressed their concerns about the application displaying the marks per page as opposed to marks per question which is what they were expecting. There were also users which felt that the displaying of error messages needed to be improved as they were often too small or stayed on the screen for too short a time.

### Fourth Iteration

Based on the feedback received during the previous iteration changes in this iteration involved changing the display of marks to conform to what users expected as well as allowing them to navigate between the marks for each question on the marking screen. The final version of this can be seen in Figure 17.

Inset final mark per question display on main marking screen

Changes were also made to the display of errors as they are now displayed using Android AlertDialogs as opposed to Android Toasts which were used before. The advantage of using AlertDialogs is that they need to be dismissed by the user thus ensuring that they are seen whereas the Toasts would appear at the bottom of the screen for a short duration. An example of each of these can be seen in Figure 18.

Insert toast vs alert dialog

In order to address the fact that crosses were being counted as ticks additional logic had to be added before the standard gesture detection. The first step in this process was to determine if the current gesture being processed was intersecting the previous one. To ensure that the application was still responsive, the intersection test was done by creating a bounding box around each gesture and checking for the intersection of the rectangles. The more computationally expensive step of determining whether or not two gestures have points in common is only performed if the bounding rectangles intersect. In order to determine if this is a cross, the two gestures should have only a single point in common. By using this method a high degree of accuracy was achieved. Although the method does not yield an accuracy of 100%, it is acceptable due to the variation which exists in handwriting and the difficulty involved in trying to accommodate all variation.

Changes to the script processing were also made which allowed for the scanning of multiple scripts at the same time. This is an effort to further decrease the amount of work required by the person responsible for scanning the documents. The final stage of script processing which involves converting the marked image files of a test script into a single PDF file, which can be emailed to the student, was also implemented.

Development of the ScriptView system was completed at the end of this iteration.

# Results

The results which had the greatest impact on the development of the application have been briefly mentioned in Section 3.9 this was done in order to provide some context as to why certain design decisions were made. The purpose of this section is to present the quantitative data collected from the user testing sessions as well as summarize some of the most common written responses. The questionnaire used can be found in Appendix A.

## First User Session

Table 1 shows the responses of all seven user from the first testing session as well as the mode and second most common response. Add web interface results

Table 1: User responses from the first user testing session

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | User 1 | User 2 | User 3 | User 4 | User 5 | User 6 | User 7 |  | Mode | Second Most Frequent |
| Question 2 | Yes | Yes | na | No | No | Yes | No |  | - | - |
| Question 3 | 4 | 3 | 4 | 5 | 3 | 4 | 4 |  | 4 | 3 |
| Question 4 | 3 | 4 | 4 | 4 | 2 | 3 | 2 |  | 4 | 2 & 3 |
| Question 5 | 3 | 2 | 2 | 2 | 1 | 5 | 3 |  | 2 | 3 |
| Question 8 | 4 | 4 | 5 | 5 | 3 | 5 | 5 |  | 5 | 4 |

Question 3 and question 8 were both related to how intuitive the interface was to use and as can be seen from the table, both received positive results. Question 4 was related to screen layout and the results were fairly distributed between effective and ineffective. As for the readability of the scripts the results were evenly divided but from discussion with the participants it was discovered that most of them found the scripts unreadable. Users expressed concern about gesture recognition and the screen layout.

## Second User Session

Table 2 shows similar data to Table 1 except that it is related to the second user session.

Table 2: User responses from the second user testing session

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | User 1 | User 2 | User 3 | User 4 | User 5 | User 6 | User 7 |  | Mode | Second Most Frequent |
| Question 2 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |  | Yes | - |
| Question 3 | 4 | 5 | 5 | 5 | 4 | 4 | 5 |  | 5 | 4 |
| Question 4 | 5 | 5 | 4 | 4 | 3 | 4 | 5 |  | 4 & 5 | 3 |
| Question 5 | 3 | 5 | 3 | 2 | 3 | 3 | 4 |  | 3 | 5, 2 & 4 |
| Question 8 | 5 | 5 | 5 | 5 | 4 | 4 | 5 |  | 5 | 4 |

The lowest response from this session was for question 4 which was related to whether or not the user felt as if the application improved their marking speed. It is worth noting that 3 of the 7 users gave varying results. In terms of script readability (question 2), all users found the script readable. In the written responses, users expressed concerns about the clarity of instructions.

## Discussion

Experience with android. Experience with tablets/stylus.

# 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.
  12. Centers for Medicare & Medicaid Services (CMS) Office of Information Service (2008). [*Selecting a development approach*](http://www.cms.gov/Research-Statistics-Data-and-Systems/CMS-Information-Technology/XLC/Downloads/SelectingDevelopmentApproach.pdf). Webarticle. United States Department of Health and Human Services (HHS). Re-validated: March 27, 2008. Retrieved 27 Oct 2008
  13. Pokkunuri, B. P. (1989). Object oriented programming.*ACM Sigplan Notices, 24*(11), 96-101.
  14. Eder, J., Kappel, G., & Schrefl, M. (1994). Coupling and cohesion in object-oriented systems.*Technical Reprot, University of Klagenfurt, Austria,*
  15. Hitz, M., & Montazeri, B. (1995). Measuring coupling and cohesion in object-oriented systems na.
  16. Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. *Journal of Extension, 50(2),* 1-5.
  17. Android design principles | android developers Retrieved from <http://developer.android.com/design/get-started/principles.html>
  18. Action bar | android developers Retrieved from <http://developer.android.com/guide/topics/ui/actionbar.html>
  19. Smith, R. (2007). An overview of the tesseract OCR engine. *Icdar, , 7* 629-633.
  20. Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: Transparency and collaboration in an open software repository. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work,*1277-1286.