USING AN ONLINE MARKING SYSTEM FOR A LARGE CLASS ENVIRONMENT

Peter M. Ostafichuk and Carol P. Jaeger
University of British Columbia
ostafichuk@mech.ubc.ca

Abstract *The use of online team marking has the potential to both simplify and expedite the process of marking exams, papers, and other artifacts. An online team marking tool (Crowdmark) has been piloted at UBC in Mechanical Engineering (125 student midterm) and two common first year introduction to engineering courses (840 student final exam, and 730 student midterm and final exam). Crowdmark, the particular software tool used, printed a unique QR code on each page of each exam and then exams were written by students in a conventional pencil-and-paper fashion. After the exam, papers were digitized and uploaded to the Crowdmark system. Following a brief training and orientation session, all marking took place by teaching assistants through the Crowdmark interface. Overall grader preference was positive, with the majority of graders expressing a strong preference for the Crowdmark system over conventional paper-based grading. In MECH 223, extensive historical data for marking time was available, and a significant reduction in marking time per exam (30%) was observed. This time savings included time saved handling papers and entering grades. Additional benefits were also observed through the use of this system: grades and histograms are available per question in real-time; time and grader tracking data is available; exam regrading is simplified; and there is a digital record of each exam for archiving purposes as well as to prevent issues of students altering papers prior to requesting regrading. Special safeguards had to be put in place due to freedom of information and privacy protection (FOIPP) requirements in British Columbia. We have observed a slightly lower cost per graded page with Crowdmark ($0.426/page) compared to a conventional exam ($0.439/page), but this includes outsourcing printing and scanning to an industrial-scale printing company. We consider this essentially cost-neutral, but like Crowdmark for all of the other benefits it offers.*

Keywords: educational technology, grading practices, assessment, analytics

1. INTRODUCTION

In engineering programs at UBC it is common for instructors to have the support of teaching assistants (TAs) for undergraduate classes. TAs are a valuable resource, and are often tasked with the grading of assignments and tests. Having this resource helps instructors to provide rapid feedback to students during the term, which supports student learning by allowing them to determine how their responses compared to full solutions provided by the instructional team. However, there are a number of challenges that instructors must address in order to ensure the integrity of the grading process. It is typical for instructors to divide tests into sections and have one individual mark all of one section across the class for the purposes of consistency. This often results in the transport and exchange of batches of test papers amongst teaching assistants. TAs are typically graduate students within the Faculty, and often have only shared office or lab space and little control over the security of these shared spaces, therefore the security of stored papers is always of concern to instructors. Supervised marking sessions for which the instructor maintains control of the papers is more secure, but less convenient in terms of scheduling efficiency, generally restricts marking times to typical business hours, and is ultimately more time consuming. The number of TA hours assigned to each course is a function of class size, but regardless of size most instructors find the number of hours insufficient to support all potentially valuable classroom or assessment activities, and therefore find it necessary to make careful decisions about where the TAs can make the biggest contribution to the successful execution of the course. Educational technology allowing online group marking, such as the Crowdmark tool (crowdmark.com) described in this paper, was felt to have the potential to save TA time such that it could be redistributed to other activities.

In the case of one of the pilot implementations of the Crowdmark online platform described in this paper, a final exam was scheduled for a class of 850 students at the end of the December examination period. Due to the late timing of the exam, most of the instructional team and TA team had plans to leave campus with insufficient time to complete the marking of the exam scripts. Though savings of time and money were the primary factors in undertaking this pilot study, flexibility of marking location turned out to be a significant consideration in the overall value of the grading tool.

In this paper, background information will be provided on the nature of two courses for which a relatively new online marking platform, Crowdmark, was tested. The Crowdmark product will then be described, with particular attention paid to the differences in exam preparation required when using this platform. Grading experience, primarily from the perspective of the instructional team and the graders are explored, along with a cost-benefit analysis. Finally, the paper concludes with a discussion of the merits and challenges associated with wider-scale adoption of the grading tool.

2. CONTEXT

Crowdmark has been piloted in two programs: a general first year introduction to engineering with 850 students (Dec 2015 to April 2016), and a second year mechanical engineering program with 125 students (April 2015). Details on each of the courses are provided below, followed by additional context related to the use in British Columbia of cloud-based systems to store student data.

2.1. MECH 223 Course

The first use of the software was for a midterm exam for MECH 223 (Mechanical Design). This course has a cohort size of approximately 125 students. This course is part of a highly integrated second year mechanical engineering program and has a core instructional team of three people along with seven teaching assistants. The highly integrated nature of the program allows for scheduling of continual assessment points across the core second year courses in the program. Rapid feedback for students in the program is essential and thus Crowdmark was piloted to assess both the potential efficiencies for the instructional team and the benefits for students. A 'traditional' delivery of a midterm exam would consist of in-house photocopying of the exam scripts. It would be typical to format a self-contained document in which the students wrote directly on the exam script rather than a question sheet and separate response booklet. Completed exams were co-graded by the instructors and teaching assistants sitting together around a table. While this ensures consistency, it adds overhead of arranging times a majority of people are available and also requires manual entry of the scores for each question into an electronic gradebook.

2.2. APSC 100 and 101 Courses

The second use of the software was in a newly designed and implemented core course in our common first year engineering program. APSC 100 (Introduction to Engineering 1) was offered for the first time in the fall of 2015. This was also the first time that the Faculty had delivered a course to the full first year cohort that involved

a full-length final exam (2.5 hours). The course was supported by a team of 10 classroom instructors, 5 laboratory/studio instructors, 4 senior graduate students acting as curriculum designers, 9 teaching assistants, and several instructional support staff members. Approximately 850 students wrote the final exam, which was scheduled on one of the last days of the December examination period. The exam consisted of a set of multiple choice questions, for which machine-readable Scantron cards were used for students to record their responses, and a significant medium and long answer portion, for which students entered their responses directly on the printed exams. The medium and long answer portion of the exam was eight pages long. The timing of the final exam and the sheer number of exam papers to be graded were both challenges faced by the instructional team.

Based on the success using Crowdmark in the APSC 100 final exam, it was used again in APSC 101, the successor course to APSC 100. There were 740 students in APSC 101, and Crowdmark was used with a 50-minute midterm exam and a 2.5-hour final exam.

2.3. Institutional Context

Due to Freedom of Information and Protection of Privacy (FOIPP) considerations, which are stringent in British Columbia, we are not able to freely use software that stores student information on non-Canadian servers. This required taking special precautions when using a cloud-based assessment tool, as will be described in Section 3.1. This also meant we did not have access to the full range of information – namely a list of email addresses for registered students – that are expected as a normal part of operating the software.

2.4. Crowdmark

Crowdmark is an online grading platform that allows multiple graders to work simultaneously from any computing device with internet access. Three key differences in using the Crowdmark grading platform as compared to traditional preparation of exams is that each exam must be custom printed to allow the insertion of a unique QR code on each page of each exam. This is accomplished by uploading a single copy of the exam to the Crowdmark web platform, upon which a document is automatically generated that contains the required number of exams, each with the unique QR code inserted on each page of each exam. The downloaded .pdf file, or files depending on total number of exams requested, must then be printed as opposed to the traditional photocopying of a single exam. In our most recent exam, with 18 pages per booklet and 800 booklets, the .pdf file had 14,400 unique pages. We have relied on professional printers to print the

exam booklets since they have the capability to do copy set stapling, whereby the single large file is automatically separated into individual stapled exam booklets.

The second key difference is that in order to match exam scripts to students, thus facilitating automatic tabulation and recording of grades, it is necessary to 'enroll' students on the Crowdmark platform. This requires the uploading of personal information for each student, or alternately an alias of some sort in the event of privacy concerns.

The third key difference is that the completed exams must be scanned and uploaded to the marking platform. While this process does represent an additional step, once complete the exams can then be graded in parallel by multiple graders, and transporting and exchanging of subsets of the papers is no longer required.

Beyond the overall time savings and flexibility in grading schedules that are expected with Crowdmark, additional benefits are derived from the archiving of original copies of completed student papers. Having long term access to the exams enhances exam integrity (e.g. a student could not modify a paper and submit it for regrading), allows retroactive mining of data not collected during the original grading of the papers (particularly useful in the support of accreditation practices), removes the time consuming process of manually entering grades into a gradebook, and facilitates the collection of data on a more granular level.

3. EXPERIENCES USING CROWDMARK

Our general experiences using Crowdmark, from initial FOIPP considerations, through grading, to returning exams are considered below.

3.1. FOIPP Considerations

Due to strict FOIPP requirements in British Columbia, special steps had to be taken to keep student data from being stored on non-Canadian servers, such as those currently used by Crowdmark. For the first exam we piloted (MECH 223, April 2015), students were assigned a unique numerical token ID during the exam period and did not include name or student number on the exam papers. A new policy adopted by the University prior to the December 2015 pilot of APSC 100 allowed us to use an opt-out procedure for recording personal data on the exams. Through announcements and electronic sign-up sheets, any students not wanting to have their name, student number, or signature stored on non-Canadian servers were able to sign up for an alias. Approximately 120 students from the 850-student class signed up for an alias, consisting of a random unique assigned alphanumeric string. During the exam, teaching assistants circulated with a list of aliases and guided each requesting student to add this to their exam. In APSC 101 (February and April

2016), the number of student requesting aliases dropped to 30 and 20, respectively, from the 740 student class. We also implemented a new procedure where students indicated on a learning management system (LMS) signup sheet that they wanted an alias, we assigned a unique easy-to-remember capital city name to them (e.g. Ottawa, Berlin, Paris), and shared this with them as a text entry in the LMS gradebook. Students were responsible to look up and remember their alias prior to the exam, and write it directly on their exam book. We brought lists of assigned aliases with us, but did not need to refer to them (i.e. everyone remembered their alias).

3.2. Grading Experiences

Grading proceeded smoothly after a short training period for the graders. Training was provided by the course instructor, and took between 30 and 60 minutes for first time users. There was no appreciable lag or delay in moving between questions or exam papers in the tool. We typically assigned each grader to a unique question, but multiple graders were able to work simultaneously. Feedback from different graders is included in the next section. After grading was complete, the anonymous papers in Crowdmark were easily and quickly matched to students (or their alias) through the exam-matching tool in the software. The resulting grades were then extracted in a .csv file with student names and IDs (or aliases) and marks for each page of the exam.

3.3. Returning Exams

Normally, student emails would be uploaded to Crowdmark during the exam creation phase, and exams would be returned directly to students via email through Crowdmark. Due to FOIPP restrictions in BC, we do not have direct access to student email addresses so exams were created using placeholder emails. Then exams were returned by sharing the unique URL that would normally be sent via email directly from Crowdmark. Since we couldn't email them, we uploaded the URLs to the LMS gradebook; the students retrieved their unique URL and copied it into a browser. With our first attempt at this in MECH 223, we placed the URLs into an alphanumeric text field in the gradebook, but due to character limits in the gradebook, the URL had to be split into three parts: a common stem and two alphanumeric strings. The students had to manually combine all three parts and paste them in their browser. In APSC 101, we used the comment field in the LMS gradebook which did not have the same restriction on number of characters. University-wide adoption of Crowdmark and subsequent integration with the LMS would resolve these issues completely, as Crowdmark would transfer data directly to the LMS.

The Crowdmark exam return time was almost instantaneous. For comparison, in APSC 100, graded papers were returned to students in person during their studio (lab/tutorial) sessions. (Leaving exam papers out in a common area for students to sort through is discouraged at UBC, in part because of privacy considerations and in part because some papers invariably go missing – strong papers are often perceived as valuable study aids.) We were able to return most papers by the end of the following week, but some teaching assistants forgot to return papers and some students were absent and unable to pick their paper up. Most of the remaining papers were returned within two weeks, but we had a steady stream of students visiting the office of the course administrator to retrieve their exams. In addition to the faster return time, the use of Crowdmark also saved 10 minutes of studio time that instead would have been used for distributing papers.

For final exams in APSC 100 and 101, exams were not returned to students, as per normal University and Faculty policy.

4. GRADER AND STUDENT RESPONSE

For each course, feedback on Crowdmark was informally solicited from graders at the end of the course. In total, there were 18 responses, with 5 each coming from MECH 223 and APSC 100, and 8 from APSC 101. The grader-reported advantages, drawbacks, and recommendations for adoption are summarized below, followed a brief discussion of the student response.

4.1. Grader-Reported Advantages

Common trends from responses to an open-ended prompt "what advantages does Crowdmark offer over conventional (paper) exams" are summarized in Table 1. The majority of graders (56%) cited the flexibility to be able to grade in any location at any time as a significant advantage of Crowdmark. The next most common comment related to the increased grading speed with Crowdmark (39% of respondents).

Table 1 - Grader Comments on Crowdmark Benefits

Grader Comments	Responses
Flexibility: can grade anywhere, anytime	10 / 18
Faster to grade	7 / 18
Can work asynchronously with other graders	3 / 18
Liked working on computer	3 / 18
Helpful to see grade distributions in real-time	2 / 18
Liked the anonymity in the papers	2 / 18

In relation to the increased speed reported above, in 2016 graders were asked to report their impression how much faster or slower they believed Crowdmark was compared to conventional grading. Graders universally reported that they felt marking with Crowdmark was faster, with estimates ranging from 10% to 70% faster (average of 37%). A more objective assessment of marking time reduction is included in Section 5.1.

4.2. Grader-Reported Drawbacks

In terms of grader perceptions of drawbacks with Crowdmark, results are summarized in Table 2. Overall, there were fewer negative comments than positive, and no strong themes. Some graders noted that they found the interface awkward or difficult to get used to, or that they writing with a stylus on a tablet was not as natural as what they were used to from other programs (such as OneNote and PowerPoint).

Table 2 - Grader Comments on Crowdmark Drawbacks

Grader Comments	Responses
Interface awkward	3 / 18
Exams written in light pencil were at times difficult to read	2 / 18
Table pen interface lag at times	2 / 18
Would like have broader selection of stamps (e.g. "½ ")	2 / 18
More difficult to give rich feedback	1 / 18
Need an internet connection	1 / 18

4.3. Grader Recommendations on Adoption

Graders in APSC 101 were asked whether they preferred the Crowdmark system or marking a conventional paper-based exam. All graders indicated a preference for Crowdmark, with 6/8 (75%) indicating a strong preference and 2/8 indicating a mild preference. Graders from all three courses were also asked to indicate whether or not they felt Crowdmark should be adopted for future use in the courses, based on a 5-point Likert scale (strongly agree to strongly disagree). The results are summarized in Figure 1 below. Overall, there is very strong grader support for the continued use of Crowdmark, with over 80% of responses mildly or strongly agreeing it should be used. Notes 1 and 2 in the figure denote graders who split their rating; 1 indicated mild agreement for the current version, but strong agreement if the pen interface could be approved, and 2 indicated mild disagreement for the current version, but mild agreement with general improvements to the user interface. (Many of the improvements sought by 2 have already been implemented in Crowdmark.)
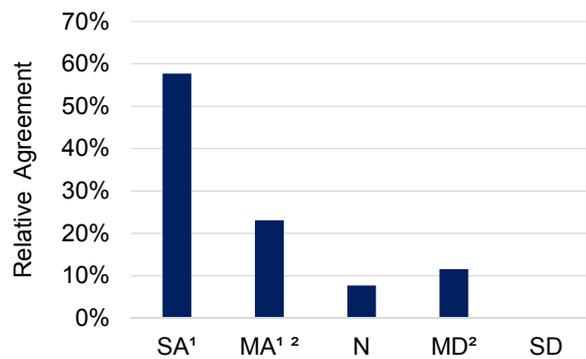
Figure 1 - Grader Agreement with Future Adoption of Crowdmark

4.4. Student Response

We did not survey students specifically about Crowdmark due to concerns of "survey burnout" (students were already completing approximate five course surveys per term in APSC 100 and 101, and three in MECH 223). No concerns about Crowdmark were raised during regular focus groups with students or with regular meetings with student representatives. Three students in MECH 223 emailed to inquire about grading, and expressed suspicion with the accuracy in marking with the new system. They were all reassured with a quick email from the course instructor.

5. COST-BENEFIT ANALYSIS

The benefits realized in using Crowdmark, and the costs incurred, are summarized in the sections below.

5.1. Benefits Realized

The reduction in marking time was identified by both course administrators and graders as a significant advantage of the Crowdmark system. Table 3 below compares the marking time using Crowdmark to historical marking of conventional paper-based exams. The exam

and question style was similar all years. For 2012 and 2013, the exam was written by individuals only. For 2014 and 2015, the exam was a two-stage format, written first by individuals and then repeated in assigned teams of 3. For the conventional exams, 4-6 markers sat together around a table and simultaneously marked in 2-3 sessions of 2-hrs each. For the Crowdmark exams, 3-4 markers (TAs and instructors) trained and calibrated together for approximately 1-hr, and then each TA marker independently marked one question. Training time on the Crowdmark system, printing time, scanning time, and data transfer time is not included with the Crowdmark numbers. Likewise, time spent generating ID tokens and anonymizing data (due to lack of integration with the LMS) is not considered as it is assumed this applies only to the pilots and would be integrated with the LMS in future, if Crowdmark is adopted.

As shown in Table 3, the average time per exam was 7.5 min with Crowdmark compared to the three year average of 10.7 min (including grades entry) with conventional marking. This represents approximately a 30% reduction in marking time. In turn, this results in a reduction in marking costs and a faster turnaround time to get graded exams back to students.

Beyond marking time, additional benefits included:
- More flexible grading for teaching assistants
- More efficient return of papers back to students
- More confidential and secure treatment of exams
- Exams are anonymous to graders
- Easier for graders to flag a paper for the instructor to review, and to easily discuss grading of a paper remotely or asynchronously
- Real-time tracking and statistical analysis of grades on a question-by-question basis
- Easier for students to request regrading of a question, and easier for the instructor to review student work
- Papers are not modifiable by students after the exam prior to a re-grading request
- Exam papers are archived for accreditation reporting

Table 3 - Summary of Marking Time for MECH 223

Year	Marking Method	Exam Format	Raw Marking Time [hrs]	Adjusted Marking Time per Exam [min]*	Total Time per Exam [min] (Including Grades Entry)
2015	Crowdmark	Two-stage	20	7.5	7.5
2014	Conventional	Two-stage	24.5	9.2	9.6
2013	Conventional	Individual only	26	11.4	11.9
2012	Conventional	Individual only	23	10.1	10.6

* Although there were more exam booklets with the two-stage exams (120 individual + 40 team = 160) than the conventional exams (120), there was a slightly higher marking load in the conventional exams due longer exam length. Each conventional

exam was approximately 14% longer than each two-stage exam. This has been accounted for in the "Adjusted Marking Time per Exam" column.

5.2. Costs Incurred

There is a time savings with Crowdmark, which translates into a direct cost savings in grader time; however, this is offset, at least in part, by additional costs for printing (since every exam booklet is now unique) and scanning (a step that is not typically done with traditional exams). Since we used a professional printing company to handle our printing and scanning, there were also courier charges. For our first pilot, with MECH 223, we had a local off-site copy shop print with copy set stapling at roughly $0.11/image (i.e. $0.22 per double-sided sheet) and we scanned pages in-house using a standard copier/scanner. We have since moved to a business-class off-site document processing company to handle both printing and scanning.

For APSC 100, the cost to print, deliver, pick-up, scan, and deliver 830 exam booklets with 20 images (10 double-sided sheets) was $2,630. The resulting total cost is roughly $0.158/image. The total time spent marking this exam between all graders (automatically tracked in Crowdmark) was 148 hours. At a nominal teaching assistant rate of $30/hour, the salary cost to mark the exam was $4,440. This gives a total cost of $7,070.

Using the previously established grading time savings of 30%, the same exam administered in a conventional format is projected to require 211 hours to grade, at a nominal cost of approximately $6,340 (an increase of $1,900). The in-house copy costs are approximately $0.05/image plus $0.01/page (2 images) for paper. For the same exam as APSC 100 (16,600 pages), this translates to a production cost of approximately $950, allowing for several hours of a staff members time to set up and monitor the copying process. This gives a resulting total cost for conventional exam of approximately $7,290.

Thus, the use of Crowdmark was just slightly better than cost-neutral. The total projected cost for production, grading, and marks entry, per single-sided exam sheet, for the two options with our APSC 100 exam was

- $0.426 for Crowdmark
- $0.439 for conventional

However, it should be noted that, as we were operating a pilot, we did not pay any usage licencing fees to Crowdmark during this time.

6. DISCUSSION AND CONCLUSIONS

Our experiences with Crowdmark have been highly favourable. We have been able to objectively demonstrate a reduction in grading time of 30%, which is consistent with graders subjective estimate of marking time reduction averaging 37%. In addition, in a large, multi-section course, exam return time was reduced by roughly a week,

a typical student. The reduction in marking time and expedited return of exams leads to more timely feedback to students, something identified throughout the educational literature as being important for learning.

We relied on an external industrial-scale company to handle printing and scanning. With this considered along with our time savings in grading, our cost per graded page for Crowdmark was roughly the same as that of a conventional paper-based exam. It is important to note that since we were piloting the software, we did not pay any licencing fees to Crowdmark, which would increase those costs.

Beyond time savings, we observed a number of significant benefits. Graders were able to work remotely, anywhere they had internet access. This was particularly important for cases such as our final exam in APSC 100, which took place one week before Christmas when many graders had already left for home. With a conventional exam, we would not have been able to complete grading until after the start of the second term in January, but with Crowdmark we essentially finalized our grades in December. In addition, graders cited liking not having to lug large boxes of exam papers around or have to arrange times to mark them in a secure location. From a teacher perspective, we appreciated the increased exam security, as well as the ability to monitor grading in real-time, and to be able to conveniently discuss marking questions with graders and with students electronically. Graders liked being able to skip questions, or flag issues for the instructor, without having to worry about setting papers aside or forgetting to come back to them. An additional benefit includes the digital archiving of exam papers for future reference, including for preventing students from editing their paper prior to a regrading request and for holding papers for accreditation purposes.

There are some limitations with the software. It currently only allows one mark to be assigned per page; if there are multiple questions on a page, the grader must tally those up before entering the grade for the page. Additionally, without integration to an LMS, precautions must be taken to satisfy FOIPP requirements. These are not debilitating issues and there are simple workarounds. Crowdmark appears to be working on or considering solutions to both.

Perhaps most convincing through our experiences was that graders indicated a strong preference for Crowdmark over conventional paper-based grading. Over 80% of graders indicated we should continue to use Crowdmark, with almost 60% strongly feeling so. Even those few graders with a preference for conventional grading saw the value of Crowdmark, and some indicated they would change their preference to Crowdmark if some additional features were added.