Making Testing Less Trying: Lessons Learned from Operating a Computer-Based Testing Facility

Craig Zilles*, Matthew West†, David Mussulman‡ and Tim Bretl§
University of Illinois at Urbana-Champaign
*Computer Science, zilles@illinois.edu
†Mechanical Science and Engineering, mwest@illinois.edu
‡Instructional Technology Facilitator, College of Engineering, mussulma@illinois.edu
§Aerospace Engineering, tbretl@illinois.edu

***Abstract*—This Innovative Practice Full Paper describes lessons learned from and operational details of a full-scale Computer-Based Testing Facility (CBTF) over a period of almost 4 years. The CBTF has grown into a key resource for enabling the graceful scaling of many of the largest classes in our College of Engineering. In Fall 2017, the CBTF served 21 courses from seven different departments and over 6,000 unique students. Over 52,000 exams were delivered, including 3,500 final exams.**

This paper discusses five main aspects of our CBTF. First, we present the basic operation of the CBTF. Second, we discuss the precautions we take to maintain a secure exam environment. Third, we discuss how we support students that require testing accommodations like extra time and/or a distraction-reduced environment. Fourth, we discuss how we organize our policies to handle exceptional circumstances with minimal intervention by faculty. Finally, we discuss the cost of operating the CBTF and how it compares to traditional exams and online services.

I. Introduction

Exams are a widely used method for summative assessment in college education, especially in introductory courses. However, at many universities, introductory courses are large (e.g., 200+ students). Running traditional pencil-and-paper exams at this scale presents management challenges that include requesting space, printing exams, proctoring, timely grading, and handling conflict exams [26], [28], [48]. These practical concerns often have more influence on how assessment is performed than pedagogical concerns.

This paper discusses the implementation of a Computer-Based Testing Facility (CBTF, Figure 1) as an alternative approach to handling exams for large classes. The goal of the CBTF is to make assessment with exams better for everyone involved—students, faculty, and course staff. Four concepts are key to achieving this goal. First, by running the exams on computers, we can write complex, authentic (e.g., numeric, programming, graphical, design) questions that are auto-gradable, allowing us to test a broad set of learning objectives with minimal grading time and providing students with immediate feedback. Second, we write question generators that use randomness to produce a collection of problems, allowing us to give each student different questions and permitting the problem generators to be used semester after semester. Third, because each student has a unique exam, we allow students to schedule their exams at a time convenient to them within a specified day range, providing students flexibility and avoiding the need to manage conflict exams. Finally, because exam scheduling and proctoring is completely handled by the CBTF, once faculty have their exam content, it is no more effort to run more frequent, smaller exams, which reduces anxiety for some students [2], [24], as well as offering second-chance exams to reduce failure rates by allowing struggling students an opportunity to review and demonstrate mastery of concepts that they missed on an exam.

Fig. 1. The Computer-based Testing Facility (CBTF) is a dedicated, proctored computer lab for summative assessment using complex, authentic exam items that permits students to flexibly schedule their exams around their other commitments.

Our CBTF is now operating in its fourth year. During that period, we have scaled from less than a thousand exams in the first semester to over 52,000 exams in Fall 2017 (Figure 2). There has been a corresponding growth in the number of courses using the CBTF (Figure 3), with most courses returning semester after semester. Furthermore, the CBTF has changed how we teach, leading to improved student learning [29] and enabling the introduction of more project and group work in large classes [44], because graduate TAs are freed from routine proctoring and grading.

This paper is intended to shed light on the operating

978-1-5386-1174-6/18/$31.00 ©2018 IEEE

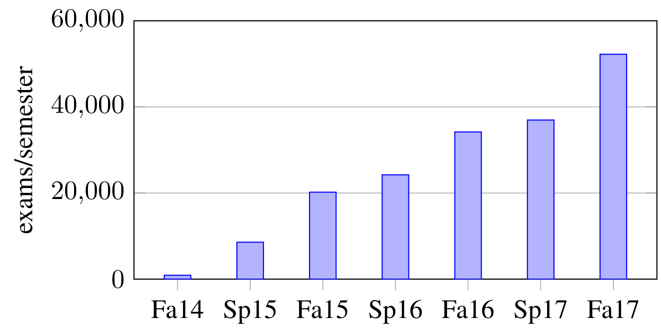

Fig. 2. The number of exams/semester in the CBTF has grown by more than a factor of 50 in the past 4 years, up to 52,224 in Fall 2017.

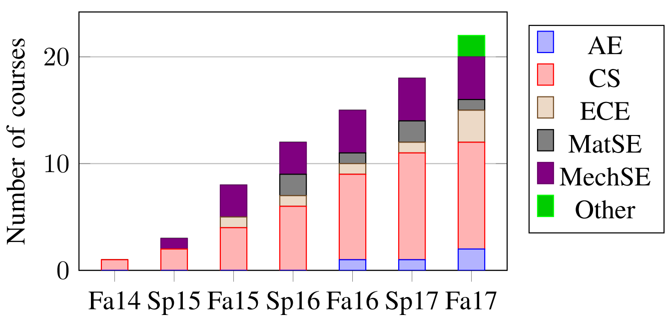

Fig. 3. The number of courses using the CBTF has grown almost linearly since its introduction. As of Fall 2017, there were seven different departments offering exams in the CBTF.

principles that have enabled this rapid adoption and substantial growth in an effort to assist others that are interested in setting up their own CBTF. It is organized as follows. We begin by discussing related work in Section II. In Section III, we present the basic operation of the CBTF. We then discuss the two most important operational issues: maintaining a secure exam environment (Section IV) and supporting students that require testing accommodations like extra time and/or a distraction-reduced environment (Section V). We then discuss how we organize our policies to handle exceptional circumstances with minimal intervention by faculty in Section VI. Finally, we discuss the cost of operating the CBTF and how it compares to traditional exams and online services (Section VII), and we conclude in Section VIII.

II. Related Work

Shacham argues that exams are the most beneficial application of computers in engineering education [40]. Two of the major benefits of computer-based testing are that it greatly reduces the overhead of running exams and they permit running exams asynchronously, allowing different students to take exams at different times [16], [22], [37], [40], [48]. In addition, the ability to provide students with immediate feedback about their errors has pedagogical value [40], and permits writing exams that allow re-trying until mastery is achieved. One major challenge of computer-based exams is generating the content [15], [23], [30].

The validity of using computers for assessment has received significant attention. While there is substantial variation in specific study results, the literature suggests that computer-based testing can be as valid as pencil-and-paper testing if sufficient precaution is taken [6]–[9], [27], [31], [46]. Prisacari and Danielson find no significant difference in the cognitive load imposed by computer or paper-based tests at the overall test level or by question type [31].

The primary threat to computer-based testing validity is test takers with complete infamiliarity with computers who can be disadvantaged on computer-based exams, perhaps resulting from anxiety impacting working memory capacity [27]. With the modern ubiquity of computers, this is likely no longer a significant concern, at least in higher education. Some older studies [42] also indicate strong aversion to computer-based testing that is generally not found in the more recent studies (e.g., [21]). Degree of familiarity with a computer can have first-order impact on test scores where that is related to the test-taking experience; for example, Russell observed that typing speed had a strong effect on scores in a language arts test (which presumably had more typing), but small and no effect on math and science tests, respectively, among middle school students [36].

Another concern with computer-based exams is the flexibility with which a student can navigate an exam. Some studies have found that tests that do not permit items to be performed out of order and reviewed before final submission can significantly affect test performance [35]. Other studies have found no score variation as a function of flexibility to change and review answers [6]. Computer-based exams appear to be less efficient to navigate non-linearly, as they were found to take, on average, 2 minutes longer for students to complete [46], and students are slower at completing exams that support non-linear navigation even when its availability doesn't affect exam scores [6]. Ackerman and Lauterman find that when under time pressure, people are less effective on computers than on paper, but have similar performance if given enough time; that is, they optimize their time more poorly on computers [1].

Research also varies somewhat about student attitudes relating to computer-based exams, but recent findings are mostly positive, especially when students are given flexibility in when to schedule their exams [37]. Student concerns include distracting keyboard noise and the inability to directly apply pencil-and-paper test-taking strategies [19]. Providing exposure to the exam format (e.g., through a practice exam) improves attitudes toward the exam [17], [32]. Studies generally find that students perceive that their scores would be the same whether the test was taken on computer or with pencil and paper [9], [19]. The biggest source of student complaints with computer-based exams are technical problems [21], [22], [40].

Computer-based programming exams in introduction-to-programming courses, often referred to as "lab exams", have received a significant amount of study as a means of ensuring that students have the skills necessary to succeed in subsequent programming courses. The earliest example seems to be the 5-

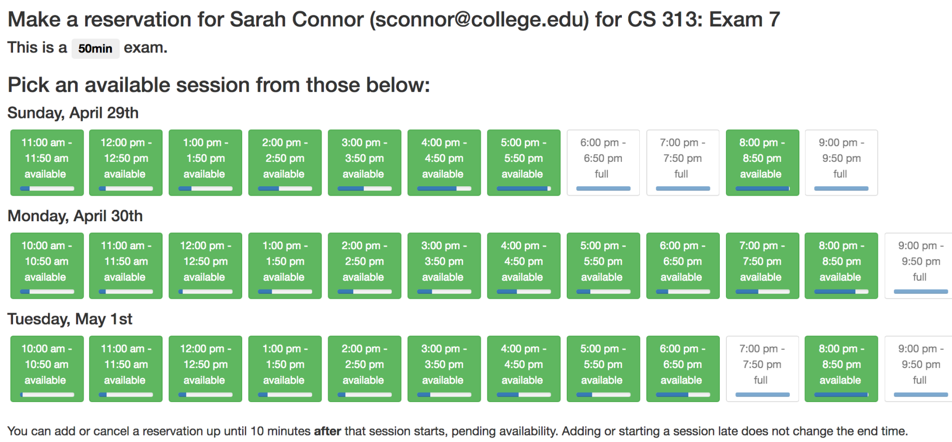

Fig. 4. Web-based interface that allows students to make reservations. Each time slot indicates by a blue bar how full it currently is.

hour proctored "mastery" exams at CMU where students were given one of a collection of similar problems to solve on a computer [11]. At first, the mastery exam fully determined the course grade, but this led students to not practice enough, so points were added back for assignments [41]. There are many similarities between this early offering and our CBTF. Thirty exam slots were offered over a two-week period to provide 1.5 times as many slots as the class would need, and students tended to bunch up in the last 4-5 slots. Students were given special accounts and all unnecessary parts of the filesystem were unmounted to prevent access to outside files [41].

Instead of replacing his traditional pencil-and-paper midterm and final, Jacobson introduced lab exams in lieu of grading programming assignments [20]. Lab exams are often managed by the course staff of a given course and are graded in bulk [3], [5], [10], [38]. Most lab-exam classes also produce multiple versions of the exam to prevent cheating from both reading off a neighbor's screen and due to the asynchronous nature of the exams. A common feature in course adoption of lab exams is the ability to re-take exams (with a different problem) at a later date, perhaps for reduced credit, when a working program cannot be produced on the first attempt [3], [20], [41]. Rajala *et al.* used the ViLLE tool to automatically grade programming exams to provide students unlimited submissions within a time limited exam [33].

Although the same kinds of logical errors are observed in both computer and pencil-and-paper programming exams, students writing code on a computer (where they can compile and test their code) make fewer errors in general [18]. This test-mode effect does not invalidate computer-based programming exams, because, as Kyllonen notes, if the real-world criterion task is more like a computer task, then the test-mode effect variance is measuring the real-world criterion variance [25]. Furthermore, lab exams have been shown to be effective. After introducing lab exams, Califf *et al.* saw a drop of the withdrawal/D/F rate in a follow-on programming course from 28.9% to 18.2% [10]. Chamillard observed better correlation between lab exam scores and written tests and the final than between traditional programming assignments and the written tests and final [12].

Some universities have developed college or campus-level resources for supporting computer-based testing. For more than ten years, the University of Helsinki has been running electronic exam rooms where students can take their final exams at a time of their choosing in a computer lab [37]. Running less than 500 exams/year on average, they grade exams manually and run exams in a lab with 16 computers without proctors, instead relying on video recording the test taker, as well as the use of multiple versions of questions. The solution at the University of Central Florida is the most like our CBTF, which seeks to support a number of large-enrollment classes with a testing center much smaller than any one class, by running the exams asynchronously [16]. They also have seen the potential that a CBTF has for reorganizing courses and benefits to student learning [43]. Others have opted for synchronous exams using student computers, in spite of the security concerns of that approach [34].

III. Standard CBTF Operation

Our CBTF implementation consists of four main components:

1) a dedicated computer lab that has been configured to serve as a secure, accommodating testing environment, as described in Sections IV and V,
2) a custom web-based scheduling tool that allows students to make exam reservations and staff to manage the operation of the CBTF,
3) software systems for the actual exam delivery (currently two different custom web-based exam tools are supported by the CBTF), and
4) a staff that includes proctors, a full-time CBTF coordinator, and technical and managerial support from our College's IT organization and two tenured faculty, which we describe in more detail in Section VII.

A typical exam experience for a student is as follows. Roughly two weeks before the exam period, a student is

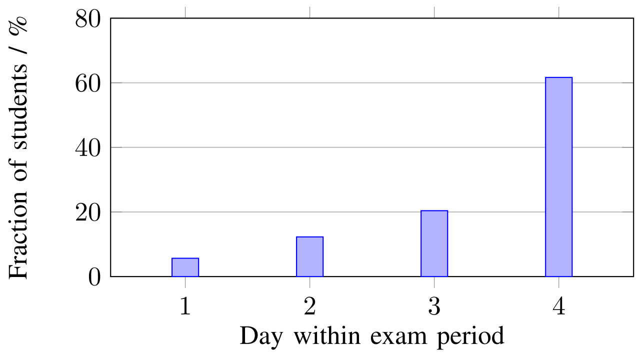

Fig. 5. The fraction of students who schedule their exams on each day of a 4-day exam period.

Fig. 6. Each CBTF workstation is numbered for seat assignment and provides a calculator and enough desk space for students to solve problems on scratch paper.

notified by email that a new exam is available for scheduling in one of their classes. Exams in the CBTF are scheduled over a period of 3-4 days, depending on the size of the class. A URL in the email brings the student to our web-based scheduler (shown in Figure 4), where they can view the available times and select one that is convenient for them. When students make a reservation, a confirmation email is sent to the student.

The CBTF runs on an hourly schedule, offering 50-minute and 1-hour-and-50-minute exams based on the needs of the class. The CBTF is open all seven days of the week from 10am to 10pm. Students can freely re-schedule their exam slot up until the beginning of that slot as space permits. This capability is heavily used, with 47% of all reservations being rescheduled. Students strongly prefer to take exams as late in the exam period as possible [49], as shown in Figure 5 for a 4-day exam.

On the day of the exam, students are sent a reminder email with the time of their scheduled exam. Students are seated for exams in the ten minutes before the exam begins. Proctors check university photo-ID cards and swipe students into the proctor interface of our scheduling software, and the student is assigned to sit at a particular computer. Students are directed to store all of their coats and bags (except a pencil or pen) on racks by the entrance before sitting at the appointed computer. Blank scratch paper is available in the CBTF and a handheld calculator is provided at every workstation, as shown in Figure 6. Students log in to the computer using their normal campus user name and then wait for a proctor to tell them to begin their exam. There is roughly 14” x 30” of desk space in front of the monitors for solving equations on scratch paper.

At this point, an exam is generated for each student by selecting items from pools of items and/or by parameterizing items (for example with different configurations or numeric constants) so that different students get different instances of the exam. Our exam delivery software is flexible enough to support any question that can be implemented in a web browser, including numeric, graphical, design, and programming questions. Students are encouraged to use the software installed on the computer to solve the exam, including Matlab for solving engineering questions and compilers/debuggers for solving programming questions. Students submit their answers and most questions are graded interactively; most exams are configured to permit multiple submissions for the same question with decreasing partial credit based on the number of attempts to reach a correct solution. During the period of the exam, there are no course staff present, only the proctors. As a result, the exam must be sufficiently self-explanatory to be interpreted by students without assistance.

During most weeks of the semester, the CBTF handles more than a dozen distinct exams each week. This means that the exam periods overlap and that students from six or more different classes may be taking an exam in the CBTF during the same time slot. The CBTF works with courses at the beginning of the semester to produce an exam schedule that meets the needs of courses without over-subscribing the available slots. We use a model to predict expected student usage for a given schedule [45] and typically aim for no more than an 85% expected utilization for any given day [49].

IV. Ensuring Exam Security

Since it is used for summative assessment, security is a paramount concern for the CBTF. Faculty and their course staff need to be able to trust the results from the exams taken in the CBTF.

Exam security begins with physical security. The CBTF is a dedicated computer lab that is used only for exam purposes and locked when not in use. When in use, the space is continually monitored by proctors. We staff with two proctors at all times to permit bathroom breaks and allow one to interact with a student while the other continues to monitor the room.

When students check into the CBTF, proctors verify their identity using the photos on their student IDs and photos available through the proctor interface of our scheduling software. The scheduling software assigns seats to students using an algorithm that attempts to put empty seats between students (if the CBTF isn’t busy) or seat students next to students taking exams from other classes. This random seat assignment makes it difficult for two students to collude during the exam period. In addition, the displays have privacy filters installed that prevent viewing the screen of adjacent computers.

Our computer workstations are configured to prevent cheating via electronic communication or access to disallowed materials. Currently, the CBTF workstations run Linux, making it relatively straightforward to turn off all networking except for a specifically white-listed set of servers, including exam delivery, authentication, and commercial software license servers. CBTF workstations have access to the same set of installed software as our public Linux lab workstations, allowing students to practice for exams in a CBTF-like software environment, but the CBTF workstations provide clean home directories using the local disk. These local home directories, which are wiped nightly, prevent students from accessing files in their normal home directories. Using Linux appears to make machine administration in the CBTF much more straightforward than previous work that used Windows [4].

As the CBTF was adopted by more courses, many students began to have multiple CBTF classes during the same semester. A few incidents where students were peeking at Exam B during a slot for which they signed up to take Exam A, led us to implement a communication channel between the CBTF scheduler and the exam delivery software that ensures that that students can only access the specified exam during a given exam reservation. With this loophole closed, we haven't observed any further attempts to subvert the digital security of the CBTF.

At present, the primary methods in which students attempt to cheat in the CBTF involve using their cell phones during the exam, bringing in unauthorized "cheat sheets", and attempting to take written notes out of the CBTF (presumably in an attempt to provide information to other students taking the exam later). As such, the CBTF is structured to control the transport of these physical objects into or out of the CBTF. Students are instructed on entry to the CBTF to place all of their belongings (e.g., coats, backpacks, cell phones) other than a writing implement on the racks that are by the entrance, as shown in Figure 7. The proctor stations (where check-in is performed) effectively partition the room into two parts, one where student things are stored and one where exams are taken. We have a zero-tolerance policy for having cell phones and other electronic devices in the exam-taking portion of the room.

While the easiest approach to eliminating cheating involving written material would be to disallow paper in the exam taking portion of the CBTF, we don't feel that is realistic yet. We believe that allowing students to use scratch paper during their exams is fundamental to both effective assessment of the students and managing their test anxiety. Instead, our approach focuses on preventing written materials from moving into or out of the exam-taking portion of the room. Most CBTF exams preclude students from bringing in written notes and instead provide any necessary formulas and documentation as a digital reference available with the exam[1]. Blank scratch paper can be picked up after a student has checked in and deposited their things on the racks and must be disposed of before leaving the exam portion of the room. We use colored paper—generally lightly-colored pastels—and change the color throughout the day to make it easier to spot paper that was brought in from outside.

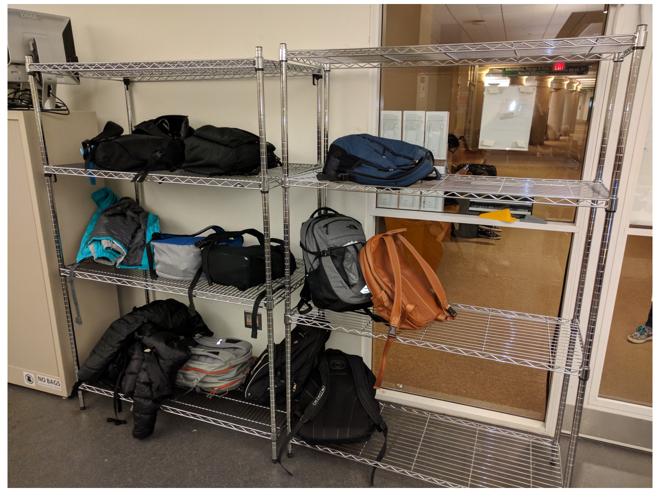

Fig. 7. The CBTF includes racks where students place their belongings (including cell phones) before proceeding to their assigned workstation.

Throughout the exam period, proctors monitor the room for cheating and suspicious behavior. The room is laid out so that proctors can walk down the aisles and from one aisle to the next on either end of the room. In addition, the room includes a number of ceiling mounted panoramic video cameras. These cameras are used both to unobtrusively monitor suspicious students while they take exams and to collect evidence for prosecuting cheating allegations. In addition, our exam delivery software allows students to anonymously report suspicious behavior of other students to the proctors.

While we do believe the level of cheating in the CBTF is as low or lower than occurs on conventional pencil-and-paper exams, we do process cheating allegations. Within our college, such allegations are handled through a web-based tool and need to be officially submitted by the faculty member teaching the course in which the academic misconduct took place. We've worked with our IT staff to provide an interface where the CBTF coordinator can prepare the allegation and the faculty member need only approve and submit the form, reducing the faculty workload.

It is important to note that these security policies are necessary but not sufficient for credible summative assessment. Because students can take the exam at different times, we have to assume that some students will seek information about the exam from others that took the exam earlier in the exam

[1]Our scheduling software does allow specific exams to be tagged as allowing students to bring in a sheet of notes and proctors can enforce that only those students bring paper into testing area. In these cases, the notes must be discarded with their scratch paper when the student leaves.

period. This means that the exam must be constructed to be resistant to this kind of communication or, in particular, to the information that can be gathered and memorized by a student seeking to exploit the exam in this way. The key strategies that we've found effective in mitigating the benefit of such an attempted exploit are: 1) randomly parameterizing items so that students need to memorize solution strategies rather than answers, 2) randomly picking items from pools, so that more items would need to be discovered and memorized, and 3) ensuring that the amount of content on the exam is significantly greater than one could hold in their short term memory. Randomizing parameters and using pool sizes of four questions has been found to be sufficient to mitigate collaborative cheating [14]. Well-constructed exams are ones where actually learning the material is a more effective strategy than trying to memorize the specifics of that exam. In general, we find that exam scores decline on average throughout the exam period, suggesting that widespread collaborative cheating isn't a problem [13].

V. Handling Testing Accommodations

The second issue that has significant influence on the design of the CBTF is supporting testing accommodations. The CBTF handles upwards of 98% of students with accommodations without additional effort on behalf of faculty and their course staff. Almost uniformly, students prefer to take their CBTF exams at the CBTF over the campus testing accommodation facility, as evidenced by the almost complete elimination of students contacting the campus center to request exam scheduling for classes using the CBTF.

Every semester, students with testing accommodations on our campus are provided a letter documenting their accommodation; we ask them to bring this letter to the CBTF so that a proctor can register their accommodation within the exam scheduling software. The two most common testing accommodations among our student population are additional time (typically a 1.5- or 2-times factor) and a "distraction-reduced" environment. Both of these are managed automatically by the scheduling software. Students requiring extra time schedule longer exam slots, but otherwise the scheduling process is normal. The CBTF provides a distraction-reduced environment through a number of workstations with partitions that block peripheral vision, as shown in Figure 8; students with such an accommodation are scheduled into that pool of seats. In addition, any student can request ear plugs.

Fig. 8. The CBTF provides distraction-reduced seats that have partitions that block peripheral vision and ear plugs are available on request.

To better support our students that use wheelchairs, the CBTF includes several workstations that are accessible for wheelchairs. These workstations are flagged as such in the seat assignment algorithm, which prioritizes such a seat assignment for those students.

Some uncommon accommodations are challenging to support without faculty involvement. One example of such an accommodation is a student with a chronic condition that can flare up causing them to miss deadlines. In such a circumstance, students are encouraged to schedule their exams early within the exam periods so they can re-schedule themselves if they have an episode. If the flare up extends beyond the testing period, the faculty member may need to be involved to grant permission to the student to take the exam outside of the testing period.

VI. Requiring Minimal Faculty Oversight

One of the key principles of the CBTF is to reduce the effort of running exams for faculty and their course staff, so that summative assessment can be performed in a manner that primarily addresses its pedagogical goals. A key element in implementing this ideal is to require minimum intervention by faculty in both the routine operation and exceptional-situation handling of CBTF exams. Specifically, we try to keep faculty informed about their exams while minimizing the amount of communication sent to and actions required from them, by isolating faculty from common exceptional situations.

Probably the most common exceptional circumstance in the CBTF is that of a missed exam. Students occasionally forget to show up to their reserved time slot and some students fall ill at the time of their exam. In a large class, even if only around 1% of your students are having such a crisis, those 5-10 students will require disproportionate course staff resources to deal with in the context of a traditional exam. The CBTF addresses missed exams with a two-pronged strategy. First, if the missed exam is sufficiently early in the exam period, the student can talk to a proctor to clear the missed exam reservation which allows the student to make another reservation within the existing exam period. If no exam slots remain that fit the student's schedule, a proctor can enable the student to schedule the exam in the one-day late period immediately following the exam period. Because this is a standard policy of the CBTF, such an extension is done without notifying the instructor, but we track all missed reservations to identify students that appear to be abusing this privilege. In practice, less than 0.5% of exams are taken during this late period.

But not all problems in the CBTF are the students' fault. Throughout the CBTF's four years of operations we've had to deal with many hardware failures and infrastructure outages. Outages that affect just one student (e.g., mouse failures, software problems on a given machine) are typically resolved by moving the student to a free computer, giving them an additional five minutes to complete their exam, and marking the machine as broken in the scheduler so that students are no longer assigned to sit at that machine. We generally only schedule 86 of the 90 seats in the CBTF so that we can gracefully tolerate a few machine failures.

We've also had a number of CBTF-wide failures, including an outage of the campus authentication servers, an outage of the campus Matlab license server, and a scheduled outage of the room's lighting about which we were not informed. In these circumstances, students' ability to complete the exam is compromised at no fault of their own. Our standard procedure is to abort the current exam attempt and allow all of the students during the affected exam slot to re-schedule their exams. In this circumstance, we will notify instructors (and/or designated TAs) of the situation and the identities of their affected students. In most cases, faculty will reset the exams for the affected students, so that they get another random instance to complete at the re-scheduled time. In general, students are quite understanding and accepting of this resolution.

The final major source of CBTF problems comes from the courses themselves in the form of problematic exams. Following too many incidents with courses deploying exams for the wrong date range, we've streamlined the deployment process to associate exam content with an exam entity in the scheduler and provide courses visual feedback that their exam is correctly bound to student reservations. This, however, doesn't prevent courses from deploying broken, confusing, or poorly designed questions. Since no course staff are present when students are taking the exam, such situations require proctors to contact the *technical contacts*—those members of the faculty and their course staff responsible for the exam—to report the problem. Because most students tend to schedule their exam towards the end of the exam period, exam bugs are generally discovered first by a small group of (generally) stronger students who made reservations early in the exam period. If these bugs can be addressed quickly, they often have only minimal impact on the class as a whole.

Exams bugs and any event that a student believes to have negatively impacted their performance on the exam can be reported to faculty via an *incident report*. An incident report includes the student's description of what happened, the proctor's report of their observations, and the actions that the proctor took to resolve the situation. The completed report is provided as information to the course instructor. The purpose of the incident reports are to validate any student issues, so that the faculty member doesn't have to rely on the student's word about some occurrence in the CBTF. The CBTF policy is that if a student leaves the CBTF without filling out an incident report then they are satisfied with the experience and have forfeited their rights to complain about the exam. As such, faculty members know that they only have to address student concerns that have associated incident reports. In general, students seem comfortable with this policy, and while some superfluous incident reports are generated, it strikes a good balance between keeping the faculty member informed and keeping communication to a minimum.

VII. Cost of the CBTF

The CBTF is only feasible because it can be operated at a reasonable cost. The cost can be broken down into three main components: facilities, staff, and consumables.

The CBTF facility is a 35' by 60' room that includes furniture, computers, and security cameras. Finding space on a college campus is always challenging, but a strong case can be made that (per square foot) our CBTF is one of the most highly utilized spaces on campus at 75% utilization 12 hours/day, seven days/week. At present, we've been most successful at acquiring space for the CBTF by converting existing computer labs, which becomes increasingly feasible as more students have their own computers and the software that they need to run for their courses can be run from their own machines. Converted computer labs have an additional advantage of being appropriately wired for power and networking to support the needs of a CBTF. Very little of the furniture in our CBTF was specifically purchased for it; it mostly came from the pre-existing computer lab and/or campus surplus. Installing the security cameras was a one-time expense of a few thousand dollars. If we had to furnish the 90 seat lab from scratch, it would probably cost around $75,000.

The primary recurring cost of the facility comes from the computers. Our IT staff refreshes computers roughly every 4-5 years, which translates to an amortized cost of around $15,000/year for new computers. In addition, some computer components need to be replaced due to wear. In the most recent semester, we replaced 4 keyboards, 7 mice, 1 network card, 9 hard drives, 1 power supply, 1 graphics card, 3 network cables, 3 power cables, and 1 complete workstation for approximately another $3,000/year in replacement parts.

Currently, the CBTF staff consists of a full-time coordinator, two full-time lead proctors, a collection of undergraduate assistant proctors, and an IT specialist that works part-time on the project. Our coordinator performs the day-to-day management of the CBTF, including the hiring, training, management, and scheduling of proctors, interfacing with faculty, and preparing any cheating allegations. In general, two proctors (one lead and one assistant) are present in the CBTF, to efficiently check students in for exams, to allow one proctor to watch the room while the other deals with student issues, and to permit breaks without leaving the room unattended. Our IT specialist handles development of the CBTF machine image and maintenance of the CBTF computer hardware and software. Staff is our largest expense at around $160,000/year.

The CBTF's only significant consumable is scratch paper. This Spring, we consumed around 170 reams of paper which translates to a little over $1,000/year. Other consumables are tissues, ear plugs, pencils, and laminating pouches, but none of these are significant expenses. In addition, hosting PrairieLearn and the CBTF scheduler on a cloud service provider costs us around $10,000/year, but is used by many classes for homework as well as exams.

As such, our annual budget for the CBTF at its current scale is around $190,000. Since we are running about 100,000 exams per year, that comes to less than $2 per exam for scheduling, proctoring, and grading. This price is the same order of magnitude as just printing a traditional mid-term or final exam and an order of magnitude cheaper than commercial exam proctoring services. Use of the CBTF has enabled some departments and courses to reconfigure their staffing to get by with fewer graduate TAs, and half of the instructors using the CBTF said that they would be prepared to accept a 20% reduction in course staff in return for use of the CBTF (13 instructors would reduce, 10 would not, 3 were neutral) [47].

One caveat that must be considered when reasoning about cost is the tendency for courses to increase the number of exams they offer when using the CBTF, for example from three 2-hour midterms (6 hours total) to semi-weekly 1-hour exams with optional second-chance exams in the off-weeks (around 9 hours total, on average). This trend is not surprising from an economic perspective, in which reducing the cost (measured in instructor time) of a desirable item (testing) makes people "buy" more of it. While this increase in testing can bring student learning gains, it also means that the 70,000 hours of CBTF exams per semester is probably replacing something more like 50,000 hours of regular exams.

VIII. Conclusion

In the past four years, our CBTF has gone from an experimental prototype implemented by a few faculty to a critical educational infrastructure in our College of Engineering. Most courses that have adopted the CBTF have no intention of ever going back to pencil-and-paper exams. The CBTF provides faculty with the potential for pedagogically better exams combined with less hassle and recurring grading time, albeit with a non-trivial up-front investment to develop exam content.

Furthermore, we don't believe that the rate of CBTF adoption on our campus will slow down any time soon. We're currently working to provision and staff a second CBTF, because our projected demand for Fall 2018 exceeds our current capacity. In addition, we're starting to get significant interest from outside of Engineering from the Statistics and Chemistry departments, and we expect that other quantitative departments like Math, Economics, and Accountancy will follow suit. Given the strong value proposition of CBTFs, we expect other colleges and universities will develop their own CBTFs, the facilitation of which is the motivation for this paper.

While we now know a lot about the basic operations of a CBTF, we believe that a lot of work remains in how to best incorporate the use of a CBTF into a course. Like any successful educational technology (e.g., projected slides, clickers) there will be good and bad ways to use a CBTF. We need to identify best practices and learn how to effectively communicate them to faculty as they adopt the CBTF. Because CBTFs free faculty from many of the constraining overheads of traditional pencil-and-paper exams, CBTFs enable a much larger space of practical summative assessment strategies that are waiting to be studied.

Acknowledgment

The authors would like to thank Mindy Fillmore and Carleen Sacris, both of whom have served as CBTF coordinator and helped develop and implement the policies described in this work. The development of the CBTF was supported initially by the Strategic Instructional Innovations Program (SIIP) of the College of Engineering at the University of Illinois, and we are grateful for the College's continued support.

References

[1] R. Ackerman and T. Lauterman, "Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure." *Computers in Human Behavior*, vol. 28, no. 5, pp. 1816–1828, 2012.

[2] J. K. Adkins and D. Linville, "Testing frequency in an introductory computer programming course," *Information Systems Education Journal*, vol. 15, no. 3, p. 22, 2017.

[3] J. P. Barros, L. Estevens, R. Dias, R. Pais, and E. Soeiro, "Using lab exams to ensure programming practice in an introductory programming course," in *Proceedings of the 8th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, (ITiCSE)*, 09 2003, pp. 16–20.

[4] E. Basar and B. Genc, "A Software System for Secure Computer Aided Exams." in *Frontiers in Education*, June 2006.

[5] J. Bennedsen and M. E. Caspersen, "Assessing process and product: A practical lab exam for an introductory programming course," *Innovation in Teaching and Learning in Information and Computer Sciences*, vol. 6, no. 4, pp. 183–202, 2007. [Online]. Available: http://www.tandfonline.com/doi/abs/10.11120/ital.2007.06040183

[6] S. M. Bodmann and D. H. Robinson, "Speed and performance differences among computer-based and paper-pencil tests," *Journal of Educational Computing Research*, vol. 31, no. 1, pp. 51–60, 2004. [Online]. Available: https://doi.org/10.2190/GRQQ-YT0F-7LKB-F033

[7] A. J. Boevé, R. R. Meijer, C. J. Albers, Y. Beetsma, and R. J. Bosker, "Introducing computer-based testing in high-stakes exams in higher education: Results of a field experiment," *PLOS ONE*, vol. 10, no. 12, pp. 1–13, 12 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0143616

[8] A. C. Bugbee Jr, "The equivalence of paper-and-pencil and computer-based testing," *Journal of Research on Computing in Education*, vol. 28, no. 3, pp. 282–299, 1996.

[9] N. Cagiltay and S. Ozalp-Yaman, "How can we get benefits of computer-based testing in engineering education?" *Computer Applications in Engineering Education*, vol. 21, no. 2, pp. 287–293, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.20470

[10] M. E. Califf and M. Goodwin, "Testing skills and knowledge: Introducing a laboratory exam in cs1," in *Proceedings of the 33rd SIGCSE Technical Symposium on Computer Science Education*, ser. SIGCSE '02. New York, NY, USA: ACM, 2002, pp. 217–221. [Online]. Available: http://doi.acm.org/10.1145/563340.563425

[11] J. Carrasquel, D. R. Goldenson, and P. L. Miller, "Competency testing in introductory computer science: the mastery examination at carnegie-mellon university," in *SIGCSE '85*, March 1985.

[12] A. T. Chamillard and K. A. Braun, "Evaluating programming ability in an introductory computer science course," in *Proceedings of the Thirty-first SIGCSE Technical Symposium on Computer Science Education*, ser. SIGCSE '00. New York, NY, USA: ACM, 2000, pp. 212–216. [Online]. Available: http://doi.acm.org/10.1145/330908.331857

[13] B. Chen, M. West, and C. Zilles, "Do performance trends suggest wide-spread collaborative cheating on asynchronous exams?" in *Learning at Scale*, 2017.

[14] ——, "How much randomization is needed to deter collaborative cheating on asynchronous exams?" in *Learning at Scale*, 2018.

[15] R. F. DeMara, B. Chen, R. Hartshorne, and R. Thripp, "Elevating participation and outcomes with computer-based assessments: An immersive development workshop for engineering faculty," *ASEE Computers in Education Journal*, vol. 8, no. 3, pp. 1–12, 2017.

[16] R. F. DeMara, N. Khoshavi, S. D. Pyle, J. Edison, R. Hartshorne, B. Chen, and M. Georgiopoulos, "Redesigning computer engineering gateway courses using a novel remediation hierarchy," in *2016 ASEE Annual Conference & Exposition*, no. 10.18260/p.26063. New Orleans, Louisiana: ASEE Conferences, June 2016, https://peer.asee.org/26063.

[17] T. Deutsch, K. Herrmann, T. Frese, and H. Sandholzer, "Implementing computer-based assessment – a web-based mock examination changes attitudes," *Computers & Education*, vol. 58, no. 4, pp. 1068 – 1075, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360131511002946

[18] S. Grissom, L. Murphy, R. McCauley, and S. Fitzgerald, "Paper vs. computer-based exams: A study of errors in recursive binary tree algorithms," in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, ser. SIGCSE '16. New York, NY, USA: ACM, 2016, pp. 6–11. [Online]. Available: http://doi.acm.org/10.1145/2839509.2844587

[19] A. Hochlehnert, K. Brass, A. Moeltner, and J. Juenger, "Does medical students' preference of test format (computer-based vs. paper-based) have an influence on performance," *BMC medical education*, vol. 11, no. 1, 2011.

[20] N. Jacobson, "Using on-computer exams to ensure beginning students' programming competency," *SIGCSE Bull.*, vol. 32, no. 4, pp. 53–56, Dec. 2000. [Online]. Available: http://doi.acm.org/10.1145/369295.369324

[21] R. James, "Tertiary student attitudes to invigilated, online summative examinations," *International Journal of Educational Technology in Higher Education*, vol. 13, no. 1, p. 19, Mar 2016. [Online]. Available: https://doi.org/10.1186/s41239-016-0015-0

[22] A. C. B. Jr. and F. M. Bernt, "Testing by computer: Findings in six years of use 1982-1988," *Journal of Research on Computing in Education*, vol. 23, no. 1, pp. 87–100, 1990. [Online]. Available: https://doi.org/10.1080/08886504.1990.10781945

[23] M. Kuikka, M. Kitola, and M.-J. Laakso, "Challenges when introducing electronic exam," *Research in Learning Technology*, vol. 22, no. 0, 2014. [Online]. Available: https://journal.alt.ac.uk/index.php/rlt/article/view/1492

[24] T. Kuo and A. Simon, "How many tests do we really need?" *College Teaching*, vol. 57, pp. 156–160, 2009.

[25] P. C. Kyllonen, "Principles for creating a computerized test battery," *Intelligence*, vol. 15, no. 1, pp. 1 – 15, 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016028969190019A

[26] E. Lee, N. Garg, C. Bygrave, J. Mahar, and V. Mishra, "Can university exams be shortened? an alternative to problematic traditional methodological approaches," in *Proceedings of the European Conference on e-Learning*. Kidmore End: Academic Conferences International Limited, 06 2015, pp. 243–250.

[27] A. S. McDonald, "The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments," *Computers & Education*, vol. 39, no. 3, pp. 299 – 312, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360131502000325

[28] R. Muldoon, "Is it time to ditch the traditional university exam?" *Higher Education Research and Development*, vol. 31, no. 2, pp. 263–265, 2012.

[29] T. Nip, E. L. Gunter, G. L. Herman, J. W. Morphew, and M. West, "Using a computer-based testing facility to improve student learning in a programming languages and compilers course," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. New York, NY, USA: ACM, 2018, pp. 568–573. [Online]. Available: http://doi.acm.org/10.1145/3159450.3159500

[30] C. G. Parshall, *Practical considerations in computer-based testing*. Springer Science & Business Media, 2002.

[31] A. A. Prisacari and J. Danielson, "Rethinking testing mode: Should i offer my next chemistry test on paper or computer?" *Computers & Education*, vol. 106, pp. 1 – 12, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360131516302317

[32] T. Rajala, E. Lokkila, R. Lindén, and M. Laakso, "Student feedback about electronic exams in introductory programming courses," in *EDULEARN15 Proceedings*, ser. 7th International Conference on Education and New Learning Technologies. IATED, July 2015, pp. 2795–2800.

[33] T. Rajala, E. Kaila, R. Lindén, E. Kurvinen, E. Lokkila, M.-J. Laakso, and T. Salakoski, "Automatically assessed electronic exams in programming courses," in *Proceedings of the Australasian Computer Science Week Multiconference*, ser. ACSW '16. New York, NY, USA: ACM, 2016, pp. 11:1–11:8. [Online]. Available: http://doi.acm.org/10.1145/2843043.2843062

[34] T. Richter and D. Boehringer, "Towards electronic exams in undergraduate engineering," in *2014 IEEE Global Engineering Education Conference (EDUCON)*, April 2014, pp. 196–201.

[35] M. Russell, A. Goldberg, and K. O'connor, "Computer-based testing and validity: a look back into the future," *Assessment in Education: Principles, Policy & Practice*, vol. 10, no. 3, pp. 279–293, 2003. [Online]. Available: https://doi.org/10.1080/0969594032000148145

[36] M. K. Russell, "Testing on computers: A follow-up study comparing performance on computer and on paper," Ph.D. dissertation, Boston College, 1999. [Online]. Available: https://www.learntechlib.org/p/127332

[37] A. Rytkönen and L. Myyry, "Student experiences on taking electronic exams at the university of helsinki," in *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2014, pp. 2114–2121.

[38] M. Saari and T. Mäkinen, "Utilizing electronic exams in programming courses: A case study," in *EDULEARN16 Proceedings : 8th International Conference on Education and New Learning Technologies*, ser. EDULEARN proceedings, July 2016, pp. 7155–7160.

[39] C. Sacris and M. Fillmore, "CBTF policies for students." [Online]. Available: https://cbtf.engr.illinois.edu/for-students/policies.html

[40] M. Shacham, "Computer-based exams in undergraduate engineering courses," *Computer Applications in Engineering Education*, vol. 6, no. 3, pp. 201–209, 1998.

[41] M. J. Stehlik and P. L. Miller, "Implementing a mastery examination in computer science," Carnegie Mellon University, Tech. Rep. CMU-CS-85-175, 1985.

[42] J. Thomas J. Ward, S. R. Hooper, and K. M. Hannafin, "The effect of computerized tests on the performance and attitudes of college students," *Journal of Educational Computing Research*, vol. 5, no. 3, pp. 327–333, 1989. [Online]. Available: https://doi.org/10.2190/4U1D-VQRM-J70D-JEQF

[43] T. Tian and R. F. DeMara, "High-fidelity digitized assessment of heat transfer fundamentals using a tiered delivery strategy," in *2018 ASEE Annual Conference & Exposition*. Salt Lake City, Utah: ASEE Conferences, June 2018.

[44] M. West, M. S. Sohn, and G. L. Herman, "Sustainable reform of an introductory mechanics course sequence driven by a community of practice," in *ASME 2015 International Mechanical Engineering Congress & Exposition (IMECE 2015)*, 2015.

[45] M. West and C. Zilles, "Modeling student scheduling preferences in a computer-based testing facility," in *Third Annual ACM Conference on Learning at Scale*, 2016, pp. 309–312.

[46] D. Zandvliet and P. Farragher, "A comparison of computer-administered and written tests," *Journal of Research on Computing in Education*, vol. 29, no. 4, pp. 423–438, 1997. [Online]. Available: https://doi.org/10.1080/08886504.1997.10782209

[47] C. Zilles, M. West, D. Mussulman, and C. Sacris, "Student and instructor experiences with a computer-based testing facility," in *10th annual International Conference on Education and New Learning Technologies (EDULEARN)*, July 2018.

[48] C. Zilles, R. T. Deloatch, J. Bailey, B. B. Khattar, W. Fagen, C. Heeren, D. Mussulman, and M. West, "Computerized testing: A vision and initial experiences," in *2015 ASEE Annual Conference & Exposition*, no. 10.18260/p.23726. Seattle, Washington: ASEE Conferences, June 2015, https://peer.asee.org/23726.

[49] C. Zilles, M. West, and D. Mussulman, "Student behavior in selecting an exam time in a computer-based testing facility," in *2016 ASEE Annual Conference & Exposition*, no. 10.18260/p.25896. New Orleans, Louisiana: ASEE Conferences, June 2016, https://peer.asee.org/25896.