Lockdown Computerized Testing Interwoven with Rapid Remediation: A Crossover Study within a Mechanical Engineering Core Course

Tian Tian
Dept. of Mechanical and Aerospace Engineering
Univeristy of Central Florida
Orlando, FL USA 32816-2450
Tian.Tian@ucf.edu

Ronald F. DeMara
Dept. of Electrical & Computer Engineering
University of Central Florida
Orlando, FL USA 32816-2362
Ronald.DeMara@ucf.edu

Su Gao
School of Teaching, Learning, and Leadership
University of Central Florida
Orlando, FL USA 32816-1250
Su.Gao@ucf.edu

***Abstract*—This paper explores the realization of viable, scalable, automated, and authentic alternatives to paper-only-based testing within Engineering disciplines. Currently, manual delivery and grading of paper-based exams incurs vast logistical burdens that have low impact to learning achievement, especially as enrollments increase. Meanwhile, Engineering's design-oriented and problem-solving emphases pose substantial challenges to the digitized delivery of assessments, and thus they warrant a substantive evaluation of their validity. To address this research need, novel Computer-Based Assessment (CBA) infrastructures and delivery protocols were launched via an IRB-approved crossover study to investigate the impact of lockdown-proctored digitized quiz and exam delivery in terms of test score validity, and learning achievement within a large-size undergraduate Mechanical and Aerospace Engineering (MAE) course.**

Results indicate that well-formed CBAs can determine scores differing as little as 0.6% from Paper-Based Assessments (PBA). Student achievement was de-correlated by technical topic of the assessment delivery mode during crossover and results revealed that the CBA delivery and remediation cohort attained up to 16.9% higher learning outcomes during summative assessment. The encouraging results are discussed in detail along with lessons learned, and suggestions for transportability of CBA approaches to other Engineering courses and institutions.

.

Keywords— Computerized testing; lock-down assessment; testing center; learning outcomes; engineering undergraduate curricula.

I. INTRODUCTION

Within large-enrollment Mechanical and Aerospace Engineering (MAE) courses, assessment delivery and grading tasks impose significant workloads. Unfortunately, diligent efforts by the instructor and Graduate Teaching Assistants (GTAs) to realize accurate paper-based assessment contribute negligible transferable progress towards those same tasks in subsequent semesters. These perpetual logistic challenges have motivated research into computerized testing, which has demonstrated some important advantages [1]. These include streamlining of the logistical overheads of exam delivery, while eliminating the time-consuming manual grading and gradebook entry tasks. Unlike paper-based testing, the effort invested to create digitized assessments is a one-time burden, which carries its benefits forward into subsequent offerings. Perpetual benefits span labor-free delivery, auto-grading, and detailed statistical analysis of assessments. These can promote the incremental improvement of question content as a means to tune assessments, engage learners, and elevate learning outcomes.

The feasibility of digitized exams within Engineering disciplines has received increasing attention in recent years [2-4]. Challenges for assessment digitization within Engineering curricula include partial credit, solution composability/traceability including handwritten work, and assessment of problem solving aspects within the constraints of contemporary Learning Management Systems (LMSs) [5]. Herein, lockdown proctored computer-based testing was evaluated as an instructional technology to realize:

1) auto-grading for formative and summative assessments,
2) secure self-paced review of solutions by students, and
3) a *Score Clarification* approach to rapid remediation utilizing a hierarchy of expertise from GTAs as tutors, with the instructor providing deeper guidance and follow-up.

Formative assessment is "a process used by teachers and students during instruction that provides feedback to adjust on-going teaching and learning to improve students' achievement of intended instructional outcomes" [6]. Formative assessment can enhance learning via the Testing Effect, and invokes aspects of a number of theories of learning, such as constructivism, behaviorism, cognitive theory, and social constructivism [7]. There are four core elements of formative assessment: identifying the "gap" between a student's status in learning and instructional goal; providing feedback to the instructor about current levels of student understanding and guiding students through their own next steps; depending on active involvement of students; and developing learning progressions toward short-term and long-term learning goals. As a pilot study, these components were designed to integrate into an undergraduate-level core required course titled *EML 4142 Fundamentals of Heat Transfer* through digitized assessments at the University of Central Florida, a large enrollment state university, during the Summer 2017 semester. For example, quizzes and auto-grading were conducted to provide feedback to both students and instructors during

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

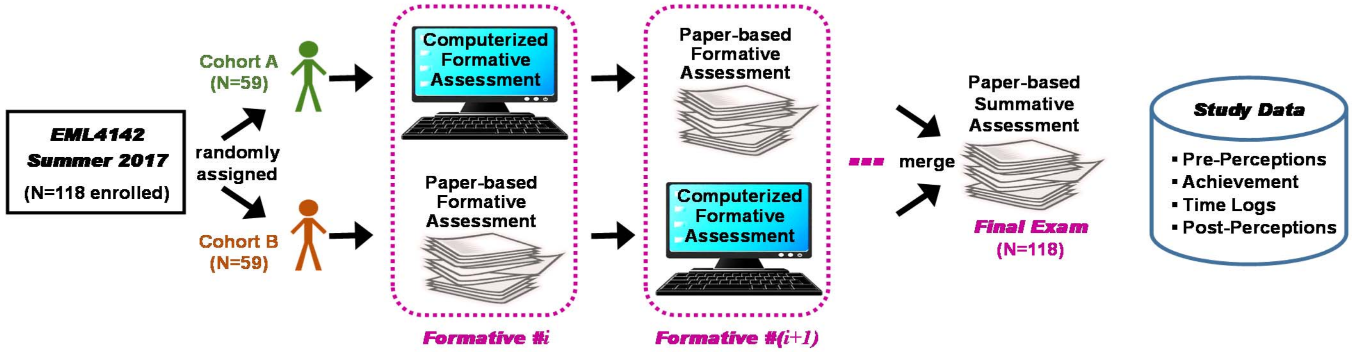


Fig. 1: Crossover Study Design. Participants are randomized to *Computerized* or *Paper-based* testing in the i^{th} *Formative Assessment*. During each successive (i^{th}+1) formative assessment, the test delivery mechanism is interchanged. All participants complete identical *Summative Assessment*. Finally, cohort-traceable achievement is analyzed. Crossover helps to maintain uniformity and equity among student participants.

specific periods of the semester. During post-quiz review, students were able to identify the learning gap through a secure self-paced review of the solutions. Then students were able to be self-motivated by a quest for partial credit to interact with GTAs and explain their problem-solving process and defend their detailed work. In this process they were able to clarify their misunderstanding, connect their prior and new learning, and enhance their learning toward long-term learning objectives. Subsequently, the instructor could also provide deeper guidance and follow-up as needed to support learning.

The use of computer-based testing via proctored lockdown delivery was evaluated using the crossover-based study design shown in Fig. 1. Under IRB approval, the crossover study randomly-partitioned students into control and intervention cohorts. The lecture and laboratory components were conducted identically for both cohorts. Meanwhile, both conventional in-class paper-based exams and computer-based exams were delivered in a mutually-exclusive flow. Computer-based exams were delivered in an *Evaluation and Proficiency Center (EPC)*, which is a College of Engineering and Computer Science (CECS)-specific testing and tutoring center [8]. The EPC resides in a vacated open computing lab and is staffed in-part by CECS GTAs who become freed up due to their abridged grading loads. EPC-based delivery allows students to complete exams in a secure manner, at a time convenient to the student, and with little burden on the part of faculty and course staff, as will be explained in detail in Section II.

Within Fig. 1 it is depicted that half of each cohort's formative assessments were delivered in the testing center via Computer-Based Assessments (CBA) and the other half of the assessments were delivered via Paper-Based Assessments (PBA). Within each cohort, the intervention of computerized delivery was interchanged with paper-based delivery during successive topic modules on a mutually-exclusive basis. At the end of the course, an identical paper-based final exam was delivered to all students. The data collected included student achievement scores on a range of formative and summative assessments. These were analyzed to investigate the topic-specific effects of digitized assessment, as will be identified within Section III. The following Research Questions (RQs) were addressed:

RQ#1: Does proctored CBA utilizing a lock-down delivery render comparable mean scores and standard deviations relative to PBA?

RQ#2: Is achievement of learning outcomes decreased when using CBA as compared to PBA?

II. Selected Related Works

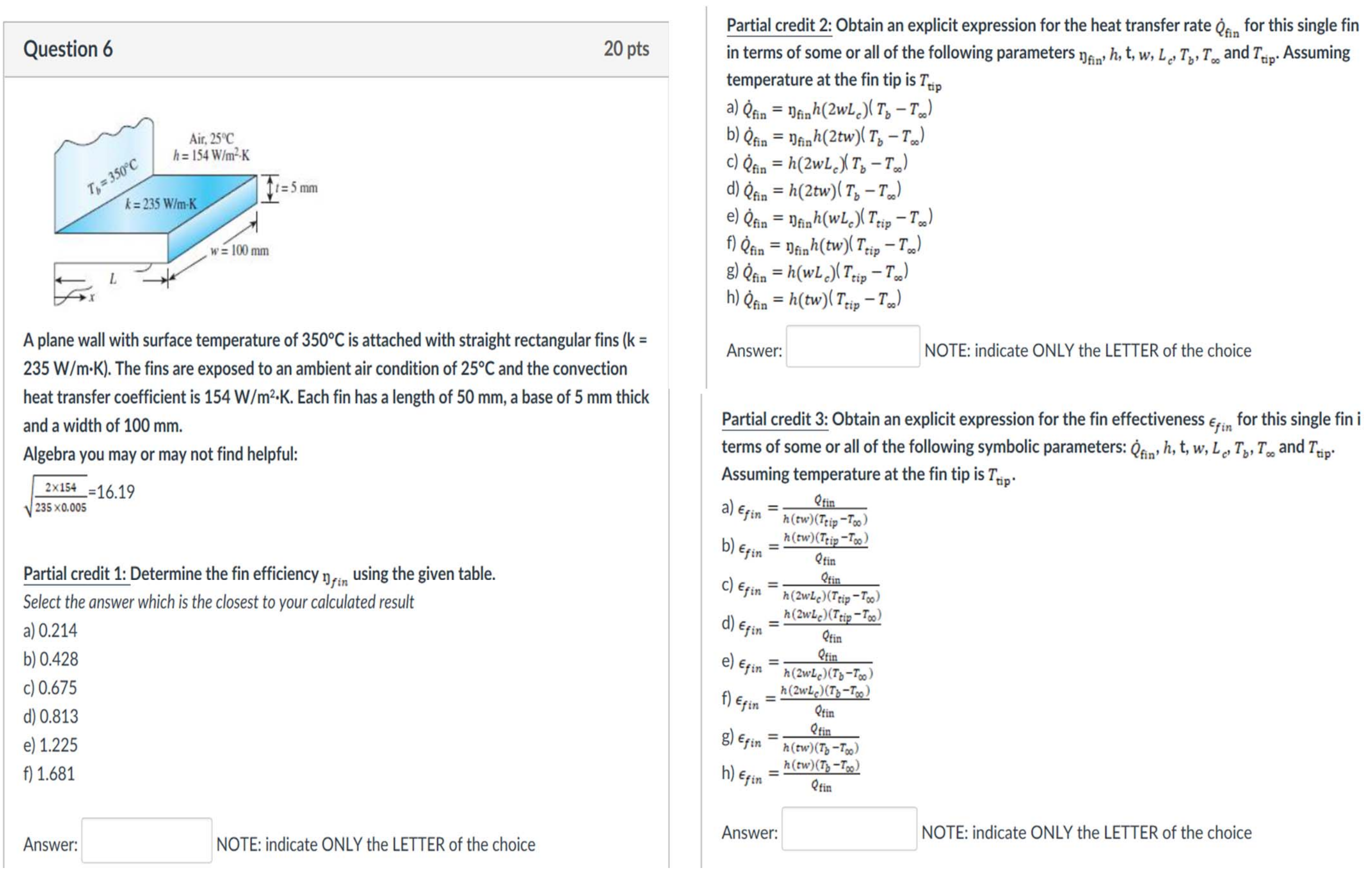
Although a comprehensive review of computerized testing would be too voluminous for the space available, some related works within Engineering curricula are highlighted herein. Clearly the trends of increasing enrollment, reduction in costs of PCs, and the success of computerized testing in other disciplines have been motivating its recent research in Engineering [2-4]. For instance, the University of Utah [3] and Brigham Young University (BYU) [9] Testing Centers (TCs) recently provide computerized testing facilities for engineering and science programs using digitized assessments. A commonality is their adoption of a secured environment overseen by proctors for evaluating significant components of the course grade. In this setting, Schurmeier et al. [3] analyzed ten years of digitized assessments from thousands of students using the University of Utah TC to elevate eight challenging topics in general chemistry. This presents an excellent example of future benefits to instructors who digitize assessments beyond the delivery mechanism itself, such as revealing trends in learners' comprehension and leveraging them for learning outcomes.

Another significant assessment digitization initiative is the Computer-Based Testing Facility (CBTF) at University of Illinois at Urbana-Champaign (UIUC) [2], which is being utilized for computer science and mechanical engineering courses. In addition to online testing and proctoring services, CBTF utilizes interactive graphical response tools for faculty to digitize their assessments. Fellin and Medicus adapted assessments within Geotechnical Engineering to variations of multiple choice formats [10]. Therein, digitization increased the achievement of undergraduate students via pre-test practice. The authors concluded that digitized assessments can realize sustained benefits that warrant the effort expended to initially compose them. While other studies have shown mixed reviews of teacher perceptions of online assessments, controlled use of CBA has shown to correlate positively with overall course grades in a 221-student controlled study in General Chemistry courses at Iowa State University [11]. On the other hand, skill inventories in a study spanning 25 sections of a college physics course showed that participation in asynchronous online CBA varied significantly based on the students' ranking within the course [12]. Thus to increase engagement and rapid remediation for all students uniformly, the testing assessment center-based approach developed herein promulgates an integrated testing and tutoring methodology [8].

Moreover, digitization enables auto-grading of assessments, which frees up graders for tutoring, a high-gain teaching and learning activity. Thus, low cost testing can lead to increased use of formative assessments to provide rapid feedback to both instructors and students, which can facilitate student learning [13, 14]. Engagement of the Testing Effect from formative assessments engages learners with retrieval practice through closed-book recall in proctored quizzes, rather than open-book efforts such as homework or online quizzes. The Testing Effect has been shown to increase learning outcomes, even for complex material [15]. Thus, proctored testing can invoke the Testing Effect via more frequent in-person digitized formative assessments, in lieu of low-gain homework submissions. This leads the learners to elevate their engagement and even ownership of the learning outcomes. Test Proctors and a lockdown browser provided high-integrity delivery of assessments to learners without Internet aides while prohibiting question archiving/multicasting to other students. A comprehensive list of services provided by the EPC infrastructure range from student services such as appointment scheduling, authentication, and stowage of unauthorized materials, pre/post-tutoring, and self-paced solution review. Faculty services span original and make-up exam delivery, proctoring, scratch paper scanning, auto-grading, gradebook entry, and video/attendance recording.

Beyond test delivery itself, rapid remediation also becomes possible within the EPC. That offers a significant new learning tool versus paper-based testing that usually incurs a week-long delay to return graded submissions. Numerous studies [16-18]

A plane wall with surface temperature of 350°C is attached with straight rectangular fins (k = 235 W/m·K). The fins are exposed to an ambient air condition of 25°C and the convection heat transfer coefficient is 154 W/m²·K. Each fin has a length of 50 mm, a base of 5 mm thick and a width of 100 mm. a) determine the fin efficiently η_{fin} using the given table. b) obtain an expression for heat transfer rate $\dot{Q}_{\text{fin}}$ for single fin in terms of η_{fin}, h, w, L_c, T_b, and T_∞. Do NOT solve with numbers. C) Obtain an explicit expression for the fin effectiveness ϵ_{fin} for this single fin in terms of some or all of the following symbolic parameters $\dot{Q}_{\text{fin}}$, h, t, w, L_c, T_b, and T_{tip}. Assuming temperature at the fin tip is T_{tip}.



Question 6 20 pts

A plane wall with surface temperature of 350°C is attached with straight rectangular fins (k = 235 W/m·K). The fins are exposed to an ambient air condition of 25°C and the convection heat transfer coefficient is 154 W/m²·K. Each fin has a length of 50 mm, a base of 5 mm thick and a width of 100 mm.

Algebra you may or may not find helpful:

$\sqrt{\frac{2\times 154}{235\times 0.005}}=16.19$

Partial credit 1: Determine the fin efficiency η_{fin} using the given table.

Select the answer which is the closest to your calculated result

a) 0.214
b) 0.428
c) 0.675
d) 0.813
e) 1.225
f) 1.681

Answer: NOTE: indicate ONLY the LETTER of the choice

Partial credit 2: Obtain an explicit expression for the heat transfer rate $\dot{Q}_{\text{fin}}$ for this single fin in terms of some or all of the following parameters η_{fin}, h, t, w, L_c, T_b, T_∞ and T_{tip}. Assuming temperature at the fin tip is T_{tip}

a) $\dot{Q}_{\text{fin}} = \eta_{\text{fin}} h(2wL_c)(T_b - T_\infty)$
b) $\dot{Q}_{\text{fin}} = \eta_{\text{fin}} h(2tw)(T_b - T_\infty)$
c) $\dot{Q}_{\text{fin}} = h(2wL_c)(T_b - T_\infty)$
d) $\dot{Q}_{\text{fin}} = h(2tw)(T_b - T_\infty)$
e) $\dot{Q}_{\text{fin}} = \eta_{\text{fin}} h(wL_c)(T_{tip} - T_\infty)$
f) $\dot{Q}_{\text{fin}} = \eta_{\text{fin}} h(tw)(T_{tip} - T_\infty)$
g) $\dot{Q}_{\text{fin}} = h(wL_c)(T_{tip} - T_\infty)$
h) $\dot{Q}_{\text{fin}} = h(tw)(T_{tip} - T_\infty)$

Answer: NOTE: indicate ONLY the LETTER of the choice

Partial credit 3: Obtain an explicit expression for the fin effectiveness ϵ_{fin} for this single fin i terms of some or all of the following symbolic parameters: $\dot{Q}_{\text{fin}}$, h, t, w, L_c, T_b, T_∞ and T_{tip}. Assuming temperature at the fin tip is T_{tip}.

a) $\epsilon_{fin} = \frac{Q_{\text{fin}}}{h(tw)(T_{tip}-T_\infty)}$
b) $\epsilon_{fin} = \frac{h(tw)(T_{tip}-T_\infty)}{Q_{\text{fin}}}$
c) $\epsilon_{fin} = \frac{Q_{\text{fin}}}{h(2wL_c)(T_{tip}-T_\infty)}$
d) $\epsilon_{fin} = \frac{h(2wL_c)(T_{tip}-T_\infty)}{Q_{\text{fin}}}$
e) $\epsilon_{fin} = \frac{Q_{\text{fin}}}{h(2wL_c)(T_b-T_\infty)}$
f) $\epsilon_{fin} = \frac{h(2wL_c)(T_b-T_\infty)}{Q_{\text{fin}}}$
g) $\epsilon_{fin} = \frac{Q_{\text{fin}}}{h(tw)(T_b-T_\infty)}$
h) $\epsilon_{fin} = \frac{h(tw)(T_b-T_\infty)}{Q_{\text{fin}}}$

Answer: NOTE: indicate ONLY the LETTER of the choice

Fig. 2. Sample free response question in a paper-based exam (above), and corresponding computer-based question (below).

Table 1: Crossover-based assessment using interchanged delivery modality in EML4142 during Summer 2017 term.

Cohort	**Quiz 1** 2017-05-25	**Midterm 1** 2017-06-08	**Quiz 2** 2017-06-22	**Midterm 2** 2017-07-06	**Quiz 3** 2017-07-20	**Quiz 4** 2017-07-27	**Final Exam** 2017-08-03
A	Computer	Paper	Paper	Computer	Computer	Paper	Paper
B	Paper	Computer	Computer	Paper	Paper	Computer	Paper

have found that timely feedback is central to guide learning. Rapid feedback allows knowledge to be assimilated promptly which additively elevates skill growth as compared to the delayed resolution of knowledge gaps [16]. Students' enthusiasm may wane when receiving feedback at the end of a subject when they perceive reduced opportunity to apply the improved understanding for gain in their course grade. [16]. This pioneers a novel *Score Clarification* technique, which self-motivates learners in a quest for partial credit to explain the problem-solving flow that they used in their formative assessment submissions. Their hand-written scratch worksheets composed during assessment are scanned-in, which elicits explanations of the solution in their own words with first-line remediation by GTA tutors, with student follow-up to the instructor. Thus, computerization of assessments can increase student engagement through in-person tutoring interwoven with assessment via Socratic discussions, fostering metacognition. During post-test review, learners conduct secure self-paced review of their formative quizzes and raise their hands to summon tutors when needed. This engages Socratic questioning which awards partial credit based on each student's scanned-in scratch work to enable remediation beyond previous uses of unscored scratch paper during CBA [19]. Thus, our integrated testing and tutoring center extends the promising aspects of an "Open Tutoring Center" where tutors are available for targeted assistance, but realizes a holistic approach, beyond project-centered assistance in "Engineering Clinics" [20].

The result is that our approach provides a uniform testing environment that realizes quiet, spacious, clean, and consistent assessment conditions. Dedicated PCs in the testing center provide stable testing machines which preclude the presence of keyloggers and malware, and trained GTAs monitor activities and camera feeds, while resolving the test-takers' concerns. Scratch paper is scanned in after students compose their free-hand design work on the blank sheets which are provided, collected, scanned-in electronically and then shredded. To enable asynchronous multiday testing windows, crosstalk between students is mitigated by disbursing questions randomly from question groups, while instantiating the values within questions randomly across each instance of exam delivery.

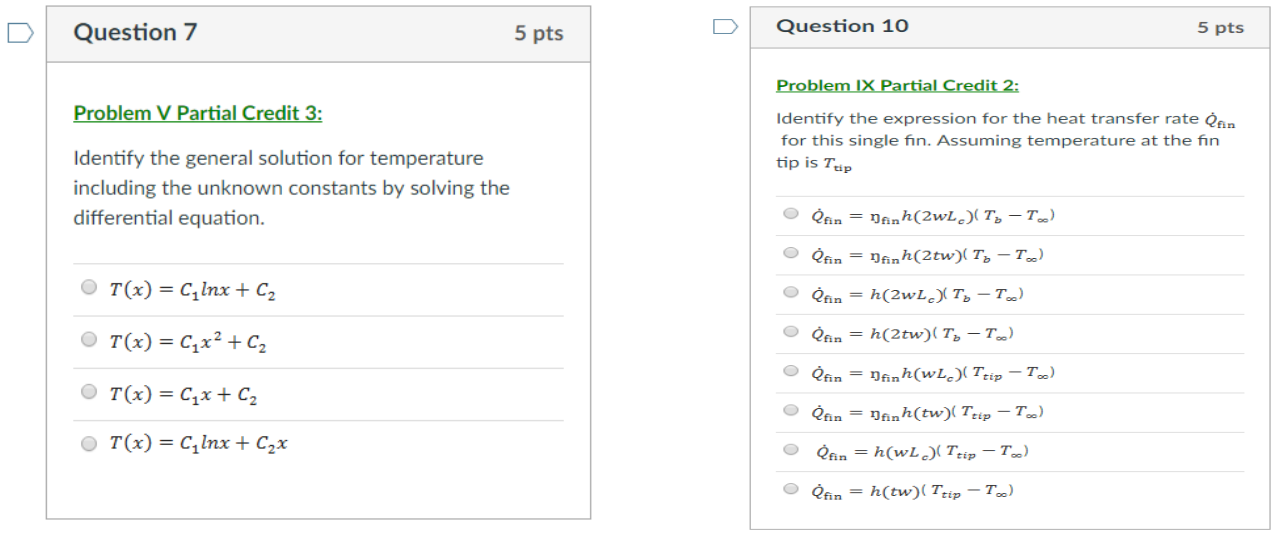
III. Research Method

As mentioned in Section I, both delivery formats were utilized in the same course in order to investigate the effectiveness of CBA relative to PBA. PBAs were delivered in the traditional classroom setting with teaching assistants serving as proctors. CBAs were undertaken in the Canvas-based Learning Management System (LMS) and delivered in the EPC testing center. Four quizzes and two midterm exams were scheduled throughout the semester. Quizzes delivered via PBA and CBA were comprised of the same questions with identical content and format. Each quiz consisted of 17 to 20 questions with a time-limit of 40 minutes enforced. Six question formats were adapted including *multiple choice*, *multiple answer*, *matching*, *true-or-false*, *fill-in-the-blank*, and *fill-in-multiple-blanks*. Quiz questions were intended to access students' understanding of fundamental concepts and principles. In contrast, the computer-based and paper-based midterms consisted of questions concerning the same topic content, although constructed in different formats. Paper-based Midterm Exam-1 and Midterm Exam-2 contained three or four free-response problems, respectively. Each free-response problem was comprised of 3-6 step-wise questions computerized for students to take in the testing center via CBA. Fig. 2 shows a sample free-response problem in PBA and its corresponding CBA counterpart.

Free-response questions in the two midterm exams using PBA required students to show their work. Partial credit was granted accordingly, whereas the graders marked those submissions manually. Free-response questions in PBA delivery were transformed to a digitized format for students to answer in the testing center via CBA. CBA submissions were graded via a hybrid machine and human-based approach. Pre-grading was initially fulfilled automatically by the Canvas LMS. Subsequently, the final score was refined by a human-driven

Table 2: Means & Standard Deviations of quizzes and exams.

	Measures	**Paper Delivery**	**Computerized Delivery**	**All Modalities**
Quiz 1	Mean	83.0%	81.9%	82.4%
	SD	12.8	13.4	13.0
	n	58	54	112
Quiz 2	Mean	79.7	80.4%	80.0%
	SD	14.1	13.2	13.6
	n	57	58	115
Quiz 3	Mean	77.9%	74.2%	76.2%
	SD	20.4	16.0	18.6
	n	61	51	112
Quiz 4	Mean	83.1%	83.4%	83.3%
	SD	15.5	14.4	14.9
	n	56	55	111
Exam 1	Mean	76.2%	84.4%	80.3%
	SD	16.9	13.0	15.6
	n	58	58	116
Exam 2	Mean	79.7%	78.1%	78.9%
	SD	15.2	17.0	16.0
	n	58	55	113



(a) (b)

Fig. 3: Sample computerized questions in (a) Midterm Exam-1 and (b) Midterm Exam-2. The question design strategy used in Midterm Exam-2 exhibits well-constructed computerized questions to help minimize scoring discrepancies when utilizing CBA.

process referred to herein as *Score Clarification*. Score Clarification self-motivates students in a quest for partial credit to explain the problem-solving flow that they used in their formative assessment submissions. Their hand-written scratch worksheets composed during assessment are scanned-in, which elicits explanations of the solution in their own words. Score Clarification is conducted via a first-line remediation by tutor GTAs, with student follow-up to the instructor as needed. The Score Clarification processes is facilitated via the extended office hours of the GTAs gained by reduced grading workloads.

Within Fig. 1 it is depicted that half of each cohort's formative assessments were delivered in the EPC via CBA and the other half of the assessments were delivered via PBA. Within each cohort, the intervention of computerized delivery was interchanged with paper-based delivery during successive topic modules on a mutually-exclusive basis. An identical paper-based Final Exam was delivered to all students.

To examine the effectiveness of CBA relative to PBA, a cross-over study as depicted in Fig. 1 randomly partitioned the class into two cohorts, with one serving as the control cohort of the other. Within each cohort, the intervention of CBA was interchanged with PBA as listed in Table 1. Cohort A took Quiz 1, Quiz 3, and Midterm Exam-2 via CBA, and Quiz 2, Quiz 4, and Midterm Exam-1 via PBA. Cohort B took Quiz 2, Quiz 4, and Midterm Exam-1 via CBA, and Quiz 1, Quiz 3, and Midterm Exam-2 via PBA. Besides the CBA intervention with test delivery, all the other activities were conducted identically for both cohorts. For example, they underwent identical lecture delivery from the same instructor, conducted the same lab components, and completed the same surveys, homework assignments, and paper-based Final Exam, etc. To answer the two research questions addressed in Section I, student test scores for the 4 quizzes and 2 midterms and student scores on specific technical topics in the summative final exam were collected and compared to address the following hypotheses:

- ***Null hypothesis I***: *There is no difference in the scoring between CBA and PBA.*
- ***Null hypothesis II:*** *There is no difference in student learning outcomes using proctored formative CBA and PBA.*

The population of the study included 118 undergraduate students enrolled in a senior level mechanical engineering course. Of the participants (aged 19-35, M=22.74, SD=3.34), 88% (*n*=104) were males and 12% (*n*=14) were females. Approximately 66% of the participants were White (*n*=78), 25% (*n*=29) Hispanic, 3% (*n*=4) Asian, 3% (*n*=3) African-American and 3% (*n*=4) multi-racial. Seniors accounted for the majority 94% (*n*=111) of the participants. A few were second degree students 5% (*n*=6) or junior degree students 1% (*n*=1). Almost all students were either Mechanical Engineering comprising 73% (*n*=86) or Aerospace Engineering majors comprising 25% (*n*=30). One student was from Digital Media (*n*=1) and one student was from Mathematics (*n*=1).

Among enrolled students (*n*=118), two students withdrew EML4142 prior to Quiz 1. Two students did not submit Midterm Exam-2, and one student was caught utilizing unauthorized materials while taking Midterm Exam-2 in the classroom via PBA. The majority (*n*=110) completed the two Midterms in both delivery modes. Three students completed both exams in the regular classroom setting. The whole class except one student (*n*=115) submitted a final exam.

IV. Study Results

The results of the research were presented in this section to answer the two research questions. In brief, students' scores on CBA can be comparable with PBA if assessments were constructed with an effective strategy. Students statistically scored consistently in both delivery modalities. The CBA cohort presented higher learning outcomes on selected technical topics in the summative assessment. A pre- and post-survey analysis revealed consistent and positive shifts in students' perception towards CBA after experiencing the intervention methods during the course. The pre- and post-survey questionnaires were made available electronically via the LMS as part of the IRB-approved study protocol. Participation was voluntary and submissions were anonymous using the LMS survey tool which enforces a limit of a single survey response per student.

In order to test the validity and reliability of the crossover study, that is to examine whether students in the two randomly-assigned cohorts possess statistical self-similarity, the four quizzes delivered via CBA and PBA were comprised of identical questions with the same content and format. Table 2 summarizes the means and standard deviations of the scores of all formative quizzes, which indicates that mean scores remained within a 4% range. These results substantiate that Cohorts A and B possess statistical self-similarity: given the same testing questions, identical performance is likely to be achieved. Besides confirming the reliability of the crossover design, this also suggests that 'computer usage anxiety' did not have a measurable impact on student performance among those participating in the study.

Significant inconsistency in Midterm Exam-1: The mean score in Midterm Exam-1 of the CBA cohort was 8.3% higher than that of the PBA cohort, which was significant. Upon detailed examination of the free-response questions in the PBA and their counterpart in the CBA, the instructor concluded that the Midterm Exam-1 via CBA was less challenging due to the lack of sufficient distracting choices to dissuade the impact of guessing. This is a challenge to note with CBA of the need to account for the impact of guessing within digitized formats. Taking one question in computerized Midterm Exam-1 as an example as shown in Fig. 3(a), the students within the CBA cohort were only required to select one option from four similar answer choices. Such clearly demands less subject matter knowledge, and thus put the CBA cohort at an advantage.

Consistent scoring in Midterm Exam-2: The mean score in Midterm Exam-2 of both cohorts were only differed by 0.6%. Fig. 3(b) shows a sample question in a computerized format within Midterm Exam-2 whereby additional answer choices were provided to enumerate a sufficient quantity of distractors to reduce the impact of guessing. Additionally, the sought-after quantity Q_{in} was expressed in terms of multiple distinct variables to obfuscate the potential for reverse engineering the correct choice while circumventing the question at hand. The authors referred to this step as *Distractor Enumeration* which can be conducted during either pre-delivery assessment design or post-delivery assessment tuning.

Table 3: Mean & Standard Deviation via Summative Assessment.

		PBA	CBA	Achievement
Heat conduction equation	Mean	70.9%	74.7%	+5.4%
	SD	4.7	4.2	
	n	57	58	
Steady heat conduction	Mean	70.8%	70.5%	-0.5%
	SD	3.8	4.2	
	n	57	58	
Heat Transfer from a finned surface	Mean	62.7%	67.0%	6.9%
	SD	2.8	3.0	
	n	58	54	
Forced convection	Mean	59.3%	69.3%	+16.9%
	SD	4.1	4.2	
	n	58	54	

As a result of Distractor Enumeration, students were less likely to "backwardly" reverse engineer the problem solution from the question statement. Similarly, it was more difficult to simply guess the correct answer by chance. The strong consistency in testing score with Midterm Exam-2 suggested that the use of a Distractor Enumeration approach can be effective to increase equivalence with PBA. It can help produce a well-composed CBA having mean scores and standard deviations comparable to PBA. Therefore, the first hypothesis regarding CBA and PBA addressed in Section IV is supported within this study based on the data collected. In summary, a vetted computer-based Engineering assessment obtains comparable score outcomes relative to PBA, thus addressing RQ#1.

To address RQ#2 regarding if proctored formative CBA can sustain learning outcomes at levels commensurate with conventional PBA, the student scores of both cohorts in the summative assessment on the four major topics assessed in the two Midterm Exams were investigated. Note the summative assessment was delivered via PBA in the traditional classroom setting and identical questions with identical format and content were assigned to all students in both cohorts. Student achievement in the summative assessment relating test delivery mode to the above-mentioned topics measured in the two

Table 4: Percentage of Students Answering Correctly versus Discrimination Index.

Question Category	Measure	CBA Cohort's Midterm Exam-1 score on each Question (Q) related to topic					
		Q1	Q2	Q3	Q4	Q5	Q6
Steady Heat Conduction	% Answered Correctly	83%	84%	83%	93%	81%	79%
	Discrimination Index	0.53	0.61	0.51	0.33	0.63	0.52
		CBA Cohort's Midterm Exam-2 score on each Question (Q) related to topic					
		Q1	Q2	Q3	Q4	Q5	Q6
Forced Convection	% Answered Correctly	62%	58%	49%	93%	76%	53%
	Discrimination Index	0.54	0.71	0.61	0.21	0.52	0.48

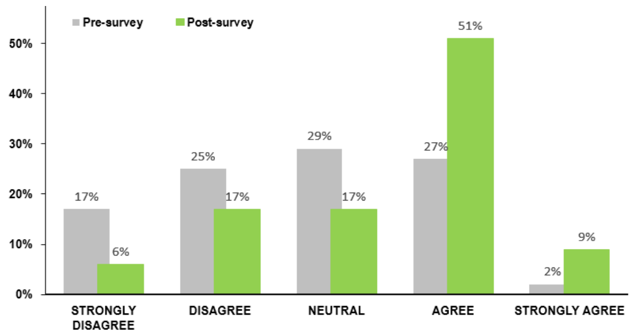


Fig. 4: Perceptions of "Regarding granting partial credit, computerized questions using step-wise question formats together with scratch paper for score clarification are effective."

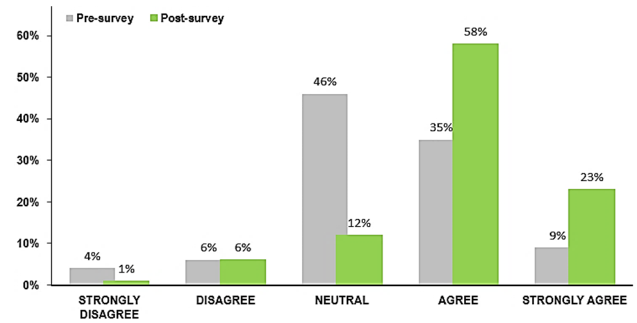


Fig. 5: Perceptions of: "Score Clarification process has contributed positively to my learning."

formative assessments was summarized in Table 3. It was revealed that the mean score of the CBA cohort was significantly higher than that of the paper-based cohort by 5.4%, 6.9%, and 16.9% on Topic #1 *Heat Conduction Equation*, Topic #2 *Heat Transfer from Finned Surfaces* and Topic #4 *External Forced Convection*, respectively.

To identify contributing factors, student scores on questions relevant to Topic 2 in the formative assessment Midterm Exam-1 delivered by CBA was examined and listed in Table 4. It was observed that most students answered those questions correctly despite their high discrimination index. This again was due to the initial challenges in question design with Midterm Exam-1, and for the same reason, students may be less motivated to visit the GTAs or the instructor for Score Clarification afterwards. Studying student scores on questions relevant to Topic 4 in the formative assessment Midterm Exam-2 via CBA as listed in Table 4, it was observed that only 49% to 58% of students answered four of the six questions correctly, which was believed having motivated students to pursue Score Clarification that had positively contributed to learning outcomes. Previously, the graded results of the PBA exams were not fully utilized by the students nor the instructional staff as an effective tool for elevating learning, but primarily as mechanism for assessment simply to assign a grade within the course. Namely, graded papers have wealth of information that can remain utilized due to lack of accessibility of the trends, lack of a direct student incentive system to learn from them, and their non-searchable paper-based format.

In this case, through the Score Clarification technique, GTAs were able to provide immediate feedback based on students' errors identified on hand-written scratch worksheets and students were able to self-monitor their understanding of specific topics and identify misconceptions. Research also shows that rapid remediation supports a more personalized approach that may appeal to some learners. It allows GTAs to instruct students individually through extra tutoring. Results in the literature have documented that rapid feedback gave students a sense of achievement while also putting students in control of their own learning [21]. Students also were engaged in articulating and defending their own conceptualizations regarding how to apply their understanding of the concepts of Heat Transfer Fundamentals in a realistic problem-solving context. This kind of expert scaffolding supports students' conceptual changes [22] and enables students to exceed the level of learning they would attain by themselves [23]. Meanwhile, there may be other conflating factors whereby CBA can lead to higher scores. Namely, digital natives may focus more attentively to computer screens than paper as observed in video records from our testing center. They may read the questions on screens more carefully and be more engaged to solve them. CBA may be more conducive to review when visibility is secured.

To gather student perception of computer-based assessment relative to paper-based assessment, an anonymous pre-survey and post-survey were administered to gather student perception towards the beginning and the end of the semester, respectively. Fig. 4 shows students' perceptions of CBA delivery of step-wise incremental question formats as a means for granting partial credit. Here, 100 and 109 out of 118 students responded to this pre- and post-survey question, respectively. In the pre-survey, only 29% Strongly Agreed or Agreed, which increased to 60% in the post-survey upon experiencing CBA methods deployed.

Fig. 5 shows learners' perceptions regarding Score Clarification. At the start of the semester, the students' preconceptions held that only 44% of respondents in the pre-survey were favorable as to whether the Score Clarification process can contribute positively to their learning. However, after being afforded the opportunity to defend their solutions using their handwritten work on the scanned-in scratch sheets, respondents' positive perceptions toward Score Clarification increased to 81% within the post-survey results.

Fig. 6 shows the student self-reported responses to sample questions related to tutoring value, which utilized the GTA time that was freed-up by auto-grading and may have contributed positively to student learning. Only a very small percentage (6%-7%) of students disagreed that the availability of teaching assistants helping with reviewing scratch paper and exams had improved their performance, increased their confidence to solve and/or communicate about engineering problems.

V. Challenges, Tradeoffs, and Transportability

Based on the quantitative and qualitative results in the previous Section, we summarize the most significant limitations facing CBA-based interventions with respect to the literature:

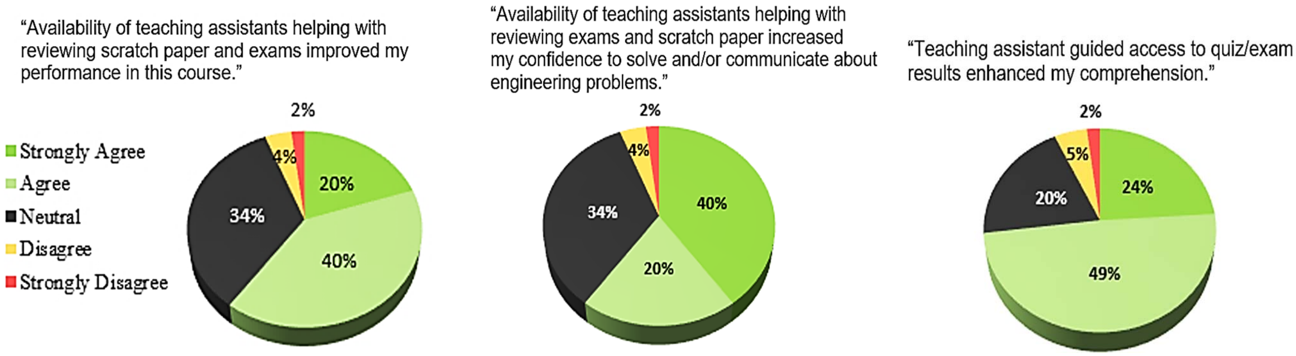


Fig. 6: Learners' responses to surveyed perceptions of tutoring value which they received from freed hours via CBA auto-grading.

- *Auto-grading support to assess partial credit* [1]: partial credit scoring limitations can be deepened from the restrictions of current LMS's quiz question formats. Although restructuring of Engineering problem statements can significantly ameliorate this concern for many topics, hybrid machine/human-based two-phase grading can be advantageous. Scanning of scratch sheets for post-test review offers a viable approach to gain some benefits from auto-grading while simultaneously retaining full flexibility for partial credit. It can also self-incentivize learners to defend their work verbally with the freed GTAs.
- *Infrastructure* [2-4][24]: reliable and secure equipment is needed to provide turnkey testing interwoven with post-test review. There are alternatives to reduce the cost and space required by the dedicated computer-based testing lab herein. The authors advocate installing lockdown browsers with an existing computer lab for use during fixed hours per week, perhaps with synchronous test delivery proctored by the course instructor. Alternatively, students may bring their own devices while utilizing commercial/freeware lockdown browsers within a WiFi-equipped classroom during days when tests are delivered.
- *Initial conversion effort* [8]: the initial digitization effort can be time-consuming for instructors. To address this hurdle, an approach utilized successfully at UCF CECS has been for the administration to underwrite the cost of an adjunct to support a one-time course release during digitization of assessments to each participating instructor.

With respect to transportability, instructors seeking to adopt CBA strategies can consider utilizing a phased approach:

1) Faculty and GTAs complete the *Digitizing and Remediating STEM Assessments* development course [8]. This is a 6-week faculty development workshop, which is made available at no cost via YouTube to the community from the authors of this manuscript.
2) Digitize only a quiz or two during the first semester,
3) Cross-verification of questions with GTAs for correctness,
4) Deliver synchronously in first term if possible, even if on paper, which avoids the need for "cloned" question variations, and
5) "Clone" the questions to reduce crosstalk among students during asynchronous testing spanning multiple days.

VI. Conclusion

Preliminary results have been sufficiently encouraging to continue the on-going cycles of additional evaluation which are needed to assess the feasibility and effectiveness of CBA within college-level Engineering courses. For instance, the techniques developed herein are being utilized in 12 courses at UCF CECS during the Summer 2018 semester which span four Engineering degree programs. Data is currently being gathered within three courses to supplement the crossover study conducted in EML4142 of these assessment mechanisms which are capable of attaining comparable mean scores, standard deviations, and score distributions relative to PBA. On-going study questions with regards to Score Clarification will focus on the feasibility to pool together the GTA personnel from multiple courses and/or degree programs using the Evaluation and Proficiency Center, with the goal of increasing the hours of availability of post-test remediation services.

References

1. Debuse, J.C. and M. Lawley, "Benefits and drawbacks of computer‐based assessment and feedback systems: Student and educator perspectives," *British Journal of Educational Technology*, 2016. **47**(2): p. 294-301.
2. Zilles, C., et al. "Computerized Testing: A Vision and Initial Experiences," in *American Society for Engineering Education Annual Conference*. 2015.
3. Schurmeier, K.D., C.G. Shepler, G.J. Lautenschlager, and C.H. Atwood, *Using Item Response Theory to Identify and Address Difficult Topics in General Chemistry*, in *Investigating Classroom Myths through Research on Teaching and Learning*. 2011, ACS Publications. p. 137-176.
4. Chen, B., R.F. DeMara, S. Salehi, and R. Hartshorne, "Elevating Learner Achievement Using Formative Electronic Lab Assessments in the Engineering Laboratory: A Viable Alternative to Weekly Lab Reports," IEEE Transactions on Education, 2017. PP(99): p. 1-10.
5. DeMara, R.F., N. Khoshavi, S. Pyle, J. Edison, R. Hartshorne, B. Chen, and M. Georgiopoulos. "Redesigning Computer Engineering Gateway Courses Using a Novel Remediation Hierarchy," in *Proceedings of American Association for Engineering Education Annual Conference*. New Orleans, LA, USA, June 26 – 29, 2016.
6. Formative Assessment for Students and Teachers (FAST), State Collaborative on Assessment and Student Standards (SCASS). (2008, October). *Attributes of effective formative assessment*. Washington, DC: Council of Chief State School Officers.
7. Moeed, A. (2015). Theorizing Formative Assessment: Time for a Change in Thinking. *The Educational Forum*, 79(2), 180–189.

8. DeMara, R.F., 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 (CoED) Journal*, 2017. 8(3).
9. Magleby, S. *Statics, CEEN103, Assessment Procedures*. Available from: https://www.physics.byu.edu/faculty/magleby/index_files/General%20index%20files/CE%20EN%20103%20Syllabus%20F2010S2.pdf accessed on 10 March 2018.
10. Fellin, W. and G. Medicus, "Multiple Choice Tests: More than a Time Saver for Teachers," *International Journal of Engineering Pedagogy*, 2015. **5**(3).
11. Prisacari, A.A. and J. Danielson, "Rethinking testing mode: Should I offer my next chemistry test on paper or computer?," *Computers & Education*, 2017. **106**: p. 1-12.
12. Nissen, J., M. Jariwala, E. Close, and B. Van Dusen, "Participation and Performance on Paper- and Computer-Based Low-Stakes Assessments." *International Journal of STEM Education*, Vol. 5. 2017.
13. Anderson, L.W., et al. "A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition." White Plains, NY: *Longman* (2001).
14. Chua, Y., "Effects of computer-based testing on test performance and testing motivation," *Comp. in Human Behavior*, 2012. **28**(5): p. 1580-1586.
15. Rawson, K.A., "The status of the testing effect for complex materials: still a winner," *Educational Psychology Review*, 2015. **27**(2): p. 327-331.
16. James, R., C. McInnis, and M. Devlin. (2002), "Assessing Learning in Australian Universities: Ideas strategies and resources for quality in student assessment," Technical Report Series, Centre for the Study of Higher Education, University of Melbourne.
17. Wiggins, G., "Feedback: How learning occurs," in Assessing Impact Evidence and Action," *American Association of Higher Education*, Bulletin #50, pp. 31-9,1997.
18. Sherman, M., Bassil, S., Lipman, D., Tuck, D., Martin, F., " Impact of auto-grading on an introductory computing course," Journal of Computing Sciences in Colleges: Volume 28 Issue 6, June 2013.
19. Prisacari, A.A. and J. Danielson, "Computer-based versus paper-based testing: Investigating testing mode with cognitive load and scratch paper use," *Computers in Human Behavior*, 2017. **77**: p. 1-10.
20. Jansson, P.M., R.P. Ramachandran, J.L. Schmalzel, and S.A. Mandayam, "Creating an agile ECE learning environment through engineering clinics," *IEEE Transactions on Education*, 2010. **53**(3): p. 455-462.
21. Yeh, S. S. (2006). "High stakes testing: Can rapid assessment reduce the pressure?" *Teachers College Record,* 108(4), 621–661.
22. Piaet, J. (1970). Science of education and psychology of the child. New York: Oxford University Press.
23. Vygotsky, L. S. (1978). Mind in society: The development of higher mental process (pp. 130-133). Cambridge, MA: Harvard University Press.
24. Nip, T., E. Gunter, G. Herman, J. Morphew, and M. West, "Using a Computer-based Testing Facility to Improve Student Learning in a Programming Languages and Compilers Course," in *Proceedings of ACM SIG on Computer Science Education*, pp. 568-573, Baltimore, Maryland, USA, February 21-24, 2018.