Article

Journal of Educational Technology Systems
0(0) 1–29

DOI: 10.1177/0047239520901863
journals.sagepub.com/home/ets
SAGE

Data Mining of Assessments to Generate Learner Remediation Teams: Method, Efficacy, and Perceptions in an Undergraduate Engineering Pilot Offering

Ronald F. DeMara[1], Salih S. Bacanli[2], Neda Bidoki[2], Jun Xu[2], Edwin Nassiff[2], Julie Donnelly[3], and Damla Turgut[2]

Abstract

This research developed an approach to integrate the complementary benefits of digitized assessments and peer learning. Its basic premise and associated hypotheses are that by using student assessments of correct and incorrect quiz answers using a fine-grained resolution to pair them into remediation peer-learning cohorts is an effective means of learning. Delivered and scored in a computer-based testing center, the assessment of digitized formative quizzes paired students whose scores indicated a different knowledge skill level so that by the end of the same week as the quiz, the

[1]Department of Electrical and Computer Engineering, University of Central Florida
[2]Department of Computer Science, University of Central Florida
[3]Karen L. Smith Faculty Center for Teaching and Learning, University of Central Florida, Orlando, FL, USA

Corresponding Author:
Ronald F. DeMara, Department of Electrical and Computer Engineering, 4328 Scorpius Street, University of Central Florida, Orlando, FL 32816–2362, USA.
Email: Ronald.Demara@ucf.edu

paired students cooperatively took an in-class quiz during a remediation session to determine if there were any learning improvements. This research article discusses this research approach and presents the findings.

Keywords
learning analytics, computer-based testing, peer teaching and learning

Motivation

Formal education of the Science, Technology, Engineering, and Mathematics (STEM) disciplines emphasizes the importance of team design, group problem-solving, and project collaboration to develop better results that are more effective. STEM, of course, includes the electrical and computer engineering (ECE) and computer science (CS) disciplines. Within the past two decades and into the foreseeable future, team design and team development skills have attained increased importance as the complexity of science and engineering problems grows significantly (Lin & Lai, 2013). This rising tide of complexity necessitates future graduates at all levels within STEM fields to function effectively as disciplinary specialists who need to work closely together and interact frequently during most phases of research and product development. While always an integral element of STEM curricula, the research emphasizes the benefits to learners immersed in collaborative learning activities to elevate the needed proficiencies of team-based skills (R. F. DeMara, Salehi, Chen, & Hartshorne, 2017). Thus, forward-looking educational technologies that demonstrate significant benefits for team-based instruction have become a high priority of researchers. From an applied standpoint, these advancements affect a broad range of STEM fields wherein lab partners, group project teams, collaborative design projects, senior design courses, and even STEM teacher participants in research experiences for teachers programs rely heavily on collaborative learning (Karbalaei, Turgut, Dagley, Vasquez, & Cho, 2018; Vasquez, Dagley, Karbalaei, Cho, & Turgut, 2019).

The research herein explores the extension of peer-learning activities by leveraging digitized formative assessments within a single ECE and CS curriculum. A promising pathway should first focus on optimizing the formation of learner teams based on data mining and statistical algorithms to structure diverse learner teams that possess complementary skills. These may include at-risk learners that can benefit from placement on teams with other students who have already mastered those complementary skills. This article investigates the use of a combination of innovative data mining and statistical analysis methods of student assessments to advance digitally mediated team building. Thus, it demonstrates a low-cost overhead and feasible approach to increase learning outcomes.

Research Objectives and Approach

Over recent years, the feasibility of digitized assessments within the ECE and CS disciplines continues to receive increasing attention (Chen, DeMara, Salehi, & Hartshorne, 2017; Schurmeier, Shepler, Lautenschlager, & Atwood, 2011; Zilles et al., 2015). Nonetheless, challenges facing digitization of assessments remain within technical curricula. For example, equitable mechanisms for partial credit assessments, grading of handwritten work, and the evaluation of creative design aspects are not available within the constraints of contemporary learning management systems (LMSs; R. F. DeMara et al., 2016). In this study, all computer-based testing occurred in a lockdown and proctored environment to enable instructional technologies to advance the following objectives:

> *Objective #1*: Conduct fair auto-grading such that the automated grading system allows instructors and graduate teaching assistants (GTAs) to quickly grade multiple assignments of formative assessments for ECE/CS course content.
>
> *Objective #2*: Generate skill-efficacy data to form peer-learning cohorts to remediate the knowledge gaps uncovered by the digitized assessments.
>
> *Objective #3*: Reallocate the time of the instructor and GTAs involved in low-gain tasks, such as grading, and redirect that time toward conducting more impactful lectures as well as recitation sessions involving remedial exercises with peer-learning cohorts.

To address Objective #1, the research study investigated the effects of both the delivery of conventional, in-class paper-based exams as well as computer-based exams. The computer-based testing occurred within the *Evaluation and Proficiency Center (EPC)* that has a specific testing and tutoring center staffed by part-time GTAs and located within the College of Engineering and Computer Science (R. F. DeMara et al., 2016). The secure EPC delivery system allows students to complete quizzes asynchronously at a time of their convenience within the center's hours of operation. This process requires minimal faculty intervention, as detailed in this section.

To address Objective #2, the formation of peer-learning cohorts included matching students with complementary knowledge gaps and skill efficacies. The formation of these automated learner cohorts involved chi-square testing and clustering analysis determined from formative student assessments over five quizzes given during the semester. Within each peer-enhanced learning cohort, students who had already acquired a particular skill became matched with those students who were deficient in the same skill, and vice versa. This approach facilitated the scalability of large enrollments and maximized opportunities for students to teach each other material that they already mastered or needed learning.

To address Objective #3, students had a 2-day window early in the week to asynchronously schedule and take a computer-based quiz in the EPC. The EPC quizzes used clone questions (multiple versions of similar test questions structured with similar settings as a standard boilerplate quiz) for each delivery instance to prevent cross talk among students. On Friday of the same week, the students were paired randomly or intelligently to conduct various remediation activities for extra credit in groups of two to four students. The formation of the intelligently clustered remediation groups used individual quiz results and paired students who incorrectly answered a portion of the quiz with students who answered that portion correctly and vice versa. As a participation incentive, students could recoup a portion of their lost quiz points by taking the collaborative remediation quiz.

The study revolved around the course entitled *COP4331: Processes for Object-Oriented Software Development*, required for all graduating seniors in ECE and CS. One section used the EPC (Efficiency and Proficiency Center) for computerized formative assessments. The test population was students enrolled in a large state university during the Fall 2017 semester and organized into a double-blinded study design as shown in Figure 1. The institutional review board (IRB) approved the randomly partitioned COP4331 students from two different sections into one control and two intervention groups. The intervention section used digitized assessments and became part of the crossover study with the control section. Group 1 consisted of the control group who used paper-based formative assessments, whereas Groups 2 and 3 consisted of the intervention group who used computerized formative assessments. All quizzes contained similar questions, and the final exam was paper-based and identical

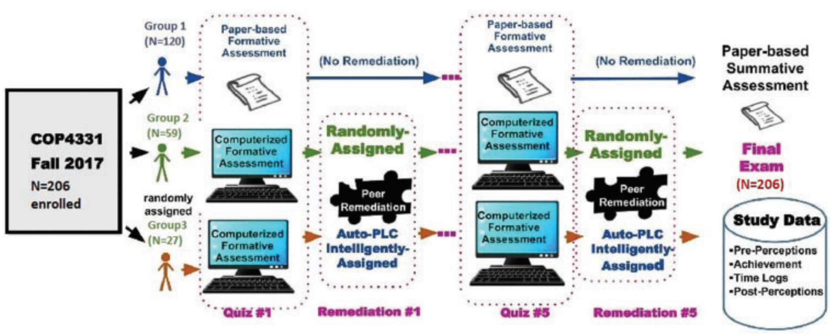

Figure 1. Double-blind study design.
PLC = peer-learning cohorts.

for all groups. Also, for Groups 2 and 3, the computerized assessments and remediation sessions were identical.

Figure 1 depicts the process of the formative assessments delivered to all three groups as either a paper-based assessment (PBA) for Group 1 or a computer-based assessment (CBA) for Groups 2 and 3. Within the latter, random selection determined if a student would fall into a randomly assigned cohort team or an intelligently assigned cohort team for peer-learning purposes. The Automated formation of Peer-Learning Cohorts (Auto-PLC) assigned members to the intelligently assigned teams.

In addition, collected data included pre/post surveys of the students' perceptions, student achievement scores from formative and summative assessments, as well as time logs of the instructors and GTAs.

Based on the aforementioned findings, the following research questions (RQs) arose:

RQ#1: Do paper-based quizzes and computerized quizzes result in comparative scores?

To answer this question, the study formed two separate sections of the course delivered by two different instructors. One was the control group using PBA with 86 students, and the other was the intervention group using CBA with 120 students. As stated earlier, both groups took a traditional paper-based final exam; however, all quiz and most final exam questions were identical in both sections; but to prevent crossover questions, many of them were cloned questions (variations of the same questions; R. F. DeMara et al., 2016).

RQ#2: Are computerized questions adequate to evaluate selected ECE and CS skills?

The formulation of cohorts in the intervention section used correlation statistics against a rudimentary skill set taxonomy. The taxonomy abstracted 20 significant elements presented in lectures as well as in the textbook and mapped into each digitized formative assessment. The randomized digitized quiz questions had multiple versions and tested all skill sets within each quiz. Questions for each skill set contained problems that challenged students to apply their engineering skills and design strategies. The data collected pointed to an affirmative answer to this RQ as evidenced by beneficial shifts between precourse and postcourse perceptions surveys collected anonymously. In all, 53 out of 86 students submitted survey responses.

RQ#3: Do students gain learning benefits, as evidenced by higher achievement on summative assessments, by participating in peer-remediation groups during recitation sessions?

We address this RQ by comparing the EPC-delivered computer-based quiz results and the postremediation assessment results within a uniform paper-based summative assessment. However, in some cases where there were no significant difference scores, there can be some assumption that students adequately acquired the content directly from the lectures or on their own. Otherwise, remediation activities did help students with low scores improve their mastery of the material. In other words, by achieving positive results after remediation, the interpretation is that students resolved their knowledge gaps as supported by their responses on the computer-based quizzes.

> *RQ#4: Do peer-learning cohorts that are formed via intelligent clustering outperform peer-learning cohorts that are formed randomly?*

To investigate this RQ, the analysis of the mean scores of postremediation assessments took place. Because the recitation sessions contained both randomly as well as intelligently clustered learning teams, as shown in Figure 1, subsequent assessments measured the differences in achievement using random matching compared with the Auto-PLC intelligently clustered peer-learning cohorts. Because a study assistant performed the analysis, it remained blind to both students and instructors. However, as discussed earlier, comparisons of student performance in the formative assessments revealed that students in the intelligently clustered groups outperformed students in the randomly clustered groups.

Selected Related Works

Digitized Assessments Within Engineering Curricula

Recent advances in testing center design and pedagogies address the support for design skills, partial credit, scanned scratch sheets, and remedial tutoring for problem-solving for ECE and CS courses referenced in the *EPC* (R. F. DeMara et al., 2016). This infrastructure promulgates an integrated testing and tutoring methodology to support a broad range of STEM programs (R.F. DeMara, Chen, Hartshorne, & Zand, 2017). Digitization enables auto-grading of assessments and alleviates grading tasks that allow greater one-on-one or group tutoring to realize high-gain teaching and learning activities. Therefore, the 120-seat capacity EPC supports low-cost assessments and enhanced remediation as well as rapid feedback. It is an effective approach for the mastery of course material and a guidepost for other instructors to follow (Anderson, Krathwohl, & Bloom, 2001). Also, the *Testing Effect* of formative assessments requires learners to recall knowledge while taking closed-book, proctored quizzes, rather than open-book efforts such as homework or online in-home quizzes. Furthermore, this effect increases learning outcomes, even for the most complex material

(Rawson, 2015). Thus, proctored testing can invoke the *Testing Effect* more frequently through in-person digitized formative assessments as learners elevate their engagement and take ownership of their learning outcomes, in place of low-gain homework submissions. Testing proctors and the use of lockdown browsers provide high-integrity delivery assessments without Internet aides and prohibit question archiving and multicasting to other students (R. F. DeMara, Salehi, & Muttineni, 2016). Services provided by the EPC infrastructure include quiz appointment scheduling, student authentication, stowage of unauthorized materials, pre/posttutoring, and self-paced solution review. Faculty services include turnkey delivery of the primary and make-up exams, proctoring, scratch paper scanning, auto-grading, gradebook entry, and video attendance recording.

Formation of Peer-Learning Cohorts as a Remediation Strategy

A two-stage exam is a well-documented method for using assessments to increase student learning rather than a single exam used for grading. The first stage of the CPA is the individual EPC assessment followed by a cohort remediation assessment. The determination of the final grade is the combination of the individual and cohort assessment scores. This method demonstrated an increase in performance for all group members including the individual performance of the higher ability group members (Heller, Keith, & Anderson, 1992; Michaelsen, Knight, & Fink, 2004). However, this only holds when groups are instructor-formed and intentionally molded to be diverse.

The characteristics of the group diversification depend on the intended outcome of the assessment. For example, if the outcome of the learning experience is to develop problem-solving skills, it is most effective to form groups that are heterogeneous in terms of their problem-solving abilities. In these groups, the high-ability members can explain concepts and strategies to low- or medium-ability members. On the other hand, low- and medium-ability members are more likely to point out simple ideas that high ability performers might overlook thus making them less efficient at solving the problem (Heller & Hollabaugh, 1992). Therefore, in ECE and CS courses, the use of students' ability levels to form effective cohort groups helps obtain positive outcomes toward necessary problem-solving skills.

Formation and management of large class student teams is a challenge. Assessing student skill sets and then grouping them based on skill levels can be chaotic when the typical size of a class is more than 100 students. However, digital tools recently introduced for other disciplines facilitate such activities (Loughry, Ohland, & Woehr, 2013). Therefore, digitally mediated team formation based on student assessment data is a necessity for effective group formation in large ECE and CS classes.

Beyond the delivery of tests, additional instructor and GTA-guided remediation also become possible when adopting digitized assessments. This includes the learning advantages resulting from the time saved from immediate CBAs versus a paper-based one, which typically requires a week or more to turn-around grades. It can also pioneer a novel *Peer-Learning Cohort* technique where self-motivated learners in search of partial credit can explain their problem-solving flow in hopes to improve their grades. To make this possible, the EPC scans student handwritten scratch worksheets used during their assessment. When a student presents him or herself to explain their solution in their own words in hopes to receive more credit, the GTA uses these scanned sheets. This first-line remediation can result in assessment regrading and provides additional teaching feedback to the student as well as the instructor. Therefore, CBAs can increase student engagement through in-person tutoring interweaved with Socratic discussions all of which foster metacognition (R. F. DeMara et al., 2017b). Figure 2 depicts learners conducting a secure self-paced review of their formative quizzes by engaging in Socratic questioning to gain partial credit based on scanned scratch worksheets. Thus, the EPC extends the promising aspects of an *Open Tutoring Center* with tutors available for targeted assistance (National Academy of Sciences, National Academy of Engineering, & Institute of Medicine, 2007) where the authors pointed out the absence of an effective, integrated, and verifiable assessment methodology. Furthermore, Auto-PLC builds on the benefits of peer learning (Watkins & Mazur, 2013) to extend the advantages of previous works (Jansson, Ramachandran, Schmalzel, & Mandayam, 2010) using learning analytics to form the peer-learning cohorts, as identified later.

Figure 2. EPC peer-learning cohort.

This project investigated the potential to find a broad vision to leverage peer learning realized through learner modeling and data mining techniques to simultaneously increase engagement, quality, creativity, and integrity beyond existing pedagogies (R. F. DeMara, Salehi, Khoshavi, Hartshorne, & Chen, 2016; Squire & Patterson, 2010; Teacher Advisory Council, 2009). This occurs during recitation and is responsive to learners' behaviors; thus, it supports interactions within the small window between instructional processes and assessment events. By leveraging student achievement data, digitally mediated adaptive team formation, and real-time monitoring to sustain instantaneous modeling of the learner, it increases the likelihood of outcomes that are highly transportable across a wide range of STEM disciplines and levels to transform the efficacy of hands-on learning. For instance, Beck (Xiong & Beck, 2014, 2015; Xiong, Wang, & Beck, 2015) and many others (Heffernan & Heffernan, 2014; Koedinger & Aleven, 2007; Razzaq & Heffernan, 2006) identified trade-offs in learning outcomes with online formative assessments through immediate feedback. When the student uses feedback, it becomes a tool for continuous growth (Epstein, Lazarus, Calvano, & Matthews, 2002; Kulik & Kulik, 1988; Webb, Stock, & McCarthy, 1994).

Table 1 summarizes related works using dynamically formed peer cohorts. Srba and Bieliková (2012) addressed the creation of dynamic short-term groups to improve the process of collaborative learning through an approach applied iteratively on a platform named PopCorm. Personal characteristics including students' knowledge, interests, and other personal characteristics are used as initial system inputs. Then, the system continuously recorded each student's previous collaboration performance the formation of future groups. Their experimental results revealed that the proposed approach outperformed two other group approaches in terms of user evaluation and feedback. Srba and Bieliková (2015) further improved their approach by creating dynamic short-term team formations, irrespective of student achievements. This approach considered a student's previous collaboration groupings and adjusted the input parameters to provide better support for their preferences during subsequent collaborations. Their experimental results showed that the study groups created by their proposed method achieved a higher collaboration quality in comparison with other groups. Henry (2013) used a novel group formation system that provided an interactive environment that allowed the instructor to experiment with different grouping parameters and algorithms in groupformation.org website. The favorable evaluations of this system involved informal surveys and discussions with students. In their experiment, most groups worked well together in completing successful semester projects and reported high satisfaction with their group learning activities.

The Auto-PLC uses a dual-pronged approach. One that interweaves digitized assessments and collaborative learning approaches to successfully remediate formative assessments and deliver knowledge refinement.

Table 1. Selected Related Works on Intelligent Grouping of Learner Teams.

Approach or tool	Application area	Group input data	Methods employed	Group setting	Outcomes
PopCorm (Srba & Bieliková, 2012)	General education domain	Personal characteristics including students' knowledge	Iterative learning students' performance	Short-term study groups	Surpasses other methods in user evaluation and feedback
Dynamic formation based on group technology (Srba & Bielikova, 2015)	General education domain	Continuous feedback from previous collaboration	Computer-based students feedback processing	Short-term study groups	The study groups achieve the higher collaboration quality in comparison with the reference groups
groupformation.org (Henry, 2013)	General education domain	Preliminary survey on students' knowledge	Best-first and partial brute force	General class group formation	Demonstrated high satisfaction within the learner team
Increasing interaction (Deibel, 2005)	Computer science education	Prior student knowledge, personality types	Encouraging interaction	Short-term in-class groups	Collaborative learning quality improves Students' attitudes toward in-class group work increase 30%
A web-based DSS assignment (Meyer, 2009)	General education domain	Prior students' knowledge	Multiple group parameters adjusting	General class project groups	Good review in both technical and user perspectives
Auto-PLC (developed herein)	Software engineering/ programming	Formative assessments	Chi-square categorization	Recitation peer learning	1.2% to 9.3% achievement increase postremediation

Note. DSS = decision support system; PLC = peer-learning cohorts.

One study proposed two-person team formations and methods for in-class group work for CS courses (Deibel, 2005). The goal was to increase interaction and promote participation among students. The first method, called the *latent jigsaw method*, assigned groups based on prior student knowledge. The second method used *Felder-Silverman learning* to promote participation by considering students' personality types. Both methods performed very well in terms of the quality of collaborative learning and students' attitudes toward in-class group work. In another study, Meyer (2009) proposed a web-based decision support system for the instructor to effectively form groups by considering different self-designated parameters. This proposed system worked in four steps: gathering preference data from students, carrying out an optimization, modifying the resulting assignment, and communicating the assignment to the students. Also, when addressing learning benefits and automation, there appears to be a significant potential for learner cohort formations to become more advanced than those mentioned earlier. The following methods follow address some new approaches.

Auto-PLC Approach

Methods Employed

The Auto-PLC, as depicted in Figure 3, incorporates the *Testing Effect* of Roediger and Karpicke (2006) by leveraging collaborative learning and incorporating collaborative learning activities (e.g., Think-Pair-Share, peer tutoring, *jigsawed* learning, etc.). According to Craven and Cooper (2016), these approaches can increase student performance in STEM courses and contribute to a learner's content knowledge while cultivating vital 21st century cross-cutting skills needed for long-term success (Craven & Cooper, 2016). Past research recognized that appropriate grouping strategies could elevate academic

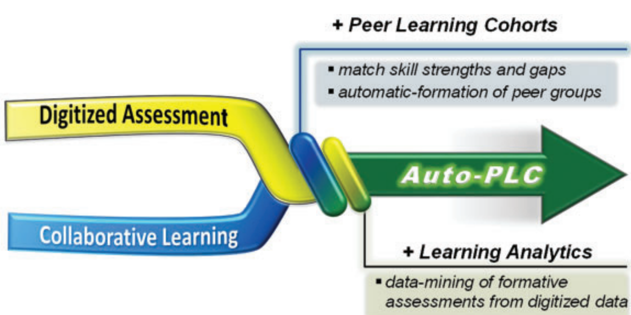

Figure 3. Auto-PLC approach.
PLC = peer-learning cohorts.

achievement within learning groups (Thomas, Bonner, Everson, & Somers, 2015). While there has been some success with computer-assisted grouping, it is known that groupings that remain static can yield mixed results (Muuro, Oboko, & Wagacha, 2016).

This study invoked a novel approach by allowing groupings to be fluid and facilitating adjustments throughout the curriculum as indicated by the learners' needs.

The matching process begins by isolating each student's quiz scores for all questions as illustrated in the sample case in Figure 4. This quiz contained seven questions in true/false, multiple-choice, and matching formats. The *students versus questions matrix* indicates the correctly answered questions for each student. In this example, each matrix column refers to a particular question, and there is a row for each of the 120 students. There is either a one or zero to indicate correctly or incorrectly answered questions, respectively. Furthermore, the seven questions were mapped onto four skill sets: program-level testing (three questions), integration testing (two questions), programming practice (one question), and box testing (one question). A mapping of skills to students in the *student versus their skills scores* matrix (columns are skills and rows are students) indicates how each student scored for each skill set. Each cell contains the maximum value obtained for each skill set, depending on the number of correctly answered questions asked for each skill set. For example, programming level testing skill cells could have a maximum value of three, corresponding to each of the three questions about this skill. Furthermore, the programming practice skill cells could have the maximum value of one. This scoring process is completed for each skill set. Using the *student versus their skills scores* matrix, the chi-square method compared students where the chi-square distance provided a number to assess the similarity of any two students. If the chi-square distance was equal to zero, this implied that the students' skills were identical. As the distance grows, the similarity between them decreases according:

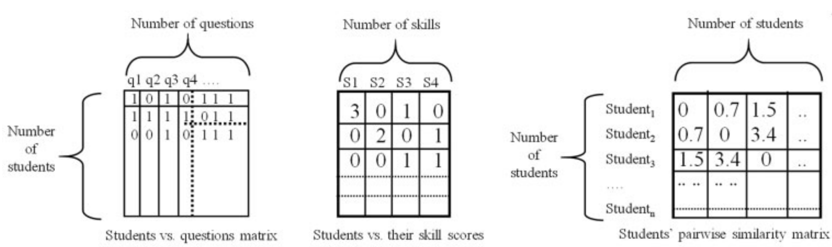

Figure 4. Intelligent clustering of students to form peer-learning cohorts using their detailed formative quiz results.

$$Chi - Square\ Distance = \sqrt{\sum \frac{(a_i - b_i)^2}{length}}$$

Also depicted in Figure 4, is the Auto-PLC constructed *students' pairwise similarity matrix* that has as many columns as rows for each of the students. Each row corresponds to a specific student, and the column value indicates the chi-square distance of that student to all other students. Auto-PLC selects a random student (row) and obtains the student (column) who has the highest chi-square score, thus selecting the student whose skill score was of the most distant. Iteratively, the process selected the farthest skill score from the remaining students until groups of two or four were formed. When placing members in a group, Auto-PLC removed all of the rows and columns of the placed members so that they would no longer be available for evaluation. The number of students placed in the cohorts changed during the semester. For example for Quiz 4, Auto-PLC created groups of four students, whereas, for Quiz 5, it created groups of two (pairs) to study the effects of group size. As outlined earlier, Auto-PLC simply identified students who differed the most in the chi-square distance and completed the pairing based solely on this criterion. The basis of diversity selection looked only at picking students who achieved high scores for a skill with those who achieved lower scores for those same skills. The process of constructing a new *students' pairwise similarity matrix* occurred for each of the five assessments to form real-time intelligent clustering (see Table 2).

Table 2. Skills Assessed in Quizzes Used to Construct the Skill Matrices Defined in Figure 4.

Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5
• Software development terms • Software engineering process roles • Development team members • Development phases • Critical path method	• UML class diagrams • Software engineering process roles • Use case diagrams	• Architecture concepts • Object-oriented programing and UML • Coupling and cohesion system-level testing	• Programming principles • Integration testing • Box testing • Program-level testing	• System-level testing • Testing team • Simulator • Testing and training documents

Note. UML = Unified Modeling Language.

Digitized Assessment and Remediation Design

To investigate the effectiveness of CBA relative to PBA assessments, each section used a different delivery format, and the results of the two sections were compared. As previously mentioned, the PBA assessments used a traditional classroom setting with TAs serving as in-class proctors. The CBA computer-based exams were delivered in the EPC, where proctoring was provided by the EPC staff who were not necessarily educated in any of the tested material. The EPC configuration used a lockdown browser to restrict Internet access and increase the integrity and security of the quizzes.

Also previously mentioned, the EPC quiz questions used formula-based formats with randomly instantiated values to help prevent cross talk among students over the 2-day examination period and used the Canvas-based LMS to deliver the quiz. There were a total of five quizzes scheduled throughout the semester as shown in Figure 5. Student progress was assessed for *Knowledge Acquisition* using the *Assessment Instrument* and finally knowledge acquisition during the *Knowledge Refinement* phase. For each quiz, of the 206 students in both sections, approximately 42% of the students underwent PBA, while the other 58% underwent the CBA. The composition of both the PBA and CBA quizzes used essentially the same questions with similar content and format. Each quiz consisted of 5 to 10 questions with a time limit of 30 minutes. The quizzes used three forms of questions: multiple choice (single and multiple answers), matching, and true and false.

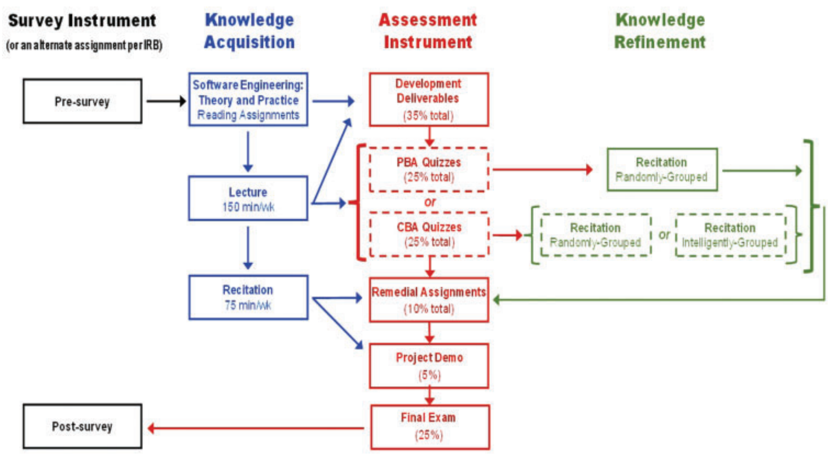

Figure 5. Course flow used in the study. Percentages indicate the weight of each activity toward the final course grade.
IRB = institutional review board; PBA = paper-based assessment; CBA = computer-based assessment.

Scoring Procedures and Implications

The five quizzes given to the students throughout the semester eliminated the need for two midterm exams as normally given in previous semesters. Again, the CBA quizzes allowed students to schedule and take them on either of two consecutive days at the beginning of the week followed by a remedial quiz given on Fridays of the same week. The remediation quizzes allowed students to earn given extra credit for correct answers. There were several issues about this format arose. A major concern was *to what benefit was it to a student who personally mastered the quiz when he/she was paired with a student who did not master the quiz?* It turned out that most students who performed well on the quiz were eager to teach their mastery of the subject to those who did not perform well. This result is consistent with the findings in previous studies in which high-performing students were able to recognize subtle mistakes in their problem-solving process when explaining it to others (Heller & Hollabaugh, 1992). Because the remediation was extra credit and not graded as a quiz, it led to an environment of little or no risk and lowered the barriers of reluctance. In addition, the remediation opportunities of those that scored better also benefited them in gaining leadership skills, reinforcing their mastery, and a feeling of accomplishment by contributing to the success of a fellow student. Besides, the higher scoring students were probably motivated by the fact that when adding in their extra credit score, a positive remediation extra credit score could make a difference between a plus or minus on their overall semester letter grade. Also, it elevated the soft skills of everyone involved in such areas as communications (needed for job interviews), problem-solving, workgroup participation, and so forth. In Delivery Results section, this report addresses these benefits. The extra credit scoring on the PBA was not possible because there were no remediation quizzes. However, it was possible to convert the majority of the CBA questions into a paper format for the PBA assessments, which reduced quiz preparation time. However, it was easier for the PBA students to receive partial credit on their quizzes because the GTA's graded them manually. These differences could have a subtle effect on the comparisons between the PBA and CBA approaches. The CBA approach employed machine- and human-based mechanisms, and the students could use scratch paper while taking the quiz. The EPC scanned and saved the scratch papers and accessed if a student requested a score clarification session with a GTA. This capability was possible because of extended GTA office hours, which were possible due to their reduced grading workloads. The LMS automatically posted the quiz grades for the CBA section.

Delivery Results

This section outlines the results of the RQs identified in the Research Objectives and Approach section. As mentioned previously, this research assessed the effectiveness of a double-blind -approved study conducted in COP4331 with 206 undergraduate students in two different sections during the Fall 2017 semester. Of that total population, 86 students comprised the control group that underwent the PBA without remediation (see Figure 5). The remaining 120 students, the intervention group, enrolled in the other section, which employed the CBA. The CBA was further divided into two subgroups labeled Group 2, composed of randomly clustered remediation groups, and Group 3, composed of intelligently clustered remediation groups. At the end of the semester, all students from both sections took an identical paper-based final exam that provided a basis of comparison.

Participant Demographics

The subgroups in the CBA were comprised of randomly clustered or intelligently clustered remediation groups and ranged in age between the age of 19 and 49 years, with a mean of 23.025 and a standard deviation of 3.82. Of the total population, males accounted for 91.7% ($n = 110$) and females 8.3% ($n = 10$). Approximately, 54.1% of the participants ($n = 65$) were White, 24.2% ($n = 29$) were Hispanic, 9.1% ($n = 11$) were African-American, 8.3% ($n = 10$) were Asian, 1.7% ($n = 2$) were multiracial, 0.8% ($n = 1$) were Hawaiian/Pacific Islanders, and 1.7% ($n = 2$) were unspecified. Undergraduate seniors accounted for the majority, that is, 89.25% ($n = 108$) of the participants, while juniors comprised 10.75% ($n = 11$). There were two second-degree seeking students, and all but one student was either Computer Engineering ($n = 44$) or Computer Science Bachelor of Science majors ($n = 75$). The one exclusion was one student who was majoring in Public Relations. Several had undergraduate research experiences (Turgut, Massi, Bacanli, & Bidoki, 2017).

Data Collection and Analysis

The study collected both quantitative and qualitative data. Quantitative data included student assessment scores, self-reported survey results, and the GTAs' time logs. Qualitative data comprised of the instructor's perceptions. The interpretation verified the quantitative analysis results. With permission, student scores from both the CBA and PBA assessments of all five quizzes along with the final exams were combined for further comparative analysis and to test each RQ. The following is a summary:

> *Learning outcomes (RQ#1): Do paper-based quizzes and computerized quizzes result in comparative scores?*

The intent was to determine whether paper-based quizzes and computer-based quizzes resulted in comparable achievement scores. Table 2 shows that the students undergoing computer-based delivery achieved higher scores on Quiz 2, while the reverse is true to a slight extent for Quiz 3. By making adjustments for the impact of guessing in the computer-based deliveries (each question had eight or more possible choices), comparable results were obtained for Quiz 4 and Quiz 5, which showed no significant differences in scores. Irrespective of the course material and delivery methods used, the evidence supports a positive answer to RQ#1.

Table 3. Quiz Scores Obtained Using Various Modes of Delivery.

Assessment	Measures	Paper-based delivery group	Computerized random group	Computerized Auto-PLC group
Quiz 1	*M*	71.5%	69.5%	70.2%
	SD	13.4%	15.4%	17.5%
	N	83	61	57
Quiz 2	*M*	46.6%	61.4%	63.1%
	SD	19.2%	14.7%	12.1%
	N	80	57	60
Quiz 3	*M*	76.1%	71.9%	71.9%
	SD	15.6%	19.6%	15.8%
	N	84	57	62
Quiz 4	*M*	81.6%	82.6%	83.8%
	SD	13.2%	14.5%	13.6%
	N	85	60	56
Quiz 5	*M*	84.1%	82.4%	84.2%
	SD	10.8%	16.5%	10.6%
	N	84	55	64
Exam 1 free response	*M*	59.5%	68.0%	
	SD	15.2%	11.6%	
Exam 2 free response	*M*	46.1%	45.7%	
	SD	11.4%	12.8%	
Final exam (paper-based)	*M*	82.9%	80.9%	
	SD	9.0%	9.4%	

Note. PLC = peer-learning cohorts.

Table 4. Digitization Equivalency—Final Exam.

Measures	Control: formative via PBA	Intervention: formative via CBA
M	82.9%	80.9%
SD	9.0%	9.4%
N	85	119

Note. PBA = paper-based assessment; CBA = computer-based assessment.

The research also studied student achievement related to the delivery assessment methods from several topics on the final exam. Tables 3 and 4 summarize these results. Students from both delivery methods statistically scored similarly on the final exam. The use of the EPC allowed the GTAs to efficiently manage their student support time, which allowed more time for student tutoring. In addition, the instructor of the CBA section was able to gain two extra lecture periods normally spent to administer the five quizzes during lecture hours.

Digitization equivalency (RQ#2): Are computerized questions adequate to evaluate selected ECE and CS skills?

This question relates to a student's ability to sufficiently refine their skills when using CBA assessments. A comparison using an identical paper-based final exam for both sections equitably compared the two sections. Table 4 indicates that the students were able to acquire the expected skill set from either of the two delivery assessment techniques. The performance on the final exam for the PBA and CBA cohorts achieved a mean score of 82.9% and 80.9%, respectively. The slight difference between the scores may be attributed to the fact that the PBA section was a daytime offering with more full-time, dedicated students compared with the other section that was offered at night. However, student scores from the PBA final exams were comparable with those in the CBA.

Remediation impact (RQ#3): Do students gain learning benefits, as evidenced by higher achievement on summative assessments, by participating in peer-remediation groups during recitation sessions?

This RQ investigated whether the students gained learning benefits by participating in peer-remediation groups during recitation sessions. The examination of learning outcomes from the quiz submission scores revealed positive outcomes especially through the review of students who fell within the lower quartile. Table 5 indicates the results of achievement on the final exam for students who received remediation compared with those who did not receive remediation. The final exam scores listed in the table reflect those who fell into the lowest 25%

Table 5. Final Exam Achievement With and Without Remediation.

Postquiz remediation activity	Mean final exam scores of those learners who achieved:				
	Lowest 25% on Quiz 1	Lowest 25% on Quiz 2	Lowest 25% on Quiz 3	Lowest 25% on Quiz 4	Lowest 25% on Quiz 5
Recipients	79.2%	80.3%	77.5%	76.8%	81.3%
Nonrecipients	77.5%	73.1%	75.0%	75.2%	72.3%
Difference	1.7%	7.2%	2.5%	1.6%	9.0%

percentile on Quizzes 1 through 5, respectively, and compared those who received the corresponding remediation following each quiz as compared with those who did not receive remediation. Using the course final exam as a uniform summative instrument, the results indicate an increase in mean final exam score of 1.7% for those who attended remediation on topics covered in Quizzes 1 through 4 and an increase up to 9.0% after remediation on topics covered in Quiz 5. This significant difference indicates that by participating in peer-grouped remediations, students saw a measurable increase in final exam scores. The outlier for Quiz 5 could be attributed to heavy semester-end workloads resulting in a smaller pool of students.

Clustering impact (RQ#4): Do peer-learning cohorts that are formed via intelligent clustering outperform peer-learning cohorts that are formed randomly?

The scores listed in Table 6 for the formative assessments indicate for all quizzes that the intelligently clustered student groups achieved higher scores by participating in peer-remediation sessions in comparison with those who were randomly clustered into groups. Although modest in gain, the data gathered for this study support the positive response to this RQ. Nonetheless, at an aggregate-level of granularity in the final exam, results listed in Table 5 further indicates an equal likelihood to achieve comparable results when students were subjected indiscriminately to either randomly clustered or intelligently clustered cohorts.

Out of the 120 students comprising both types of cohort formulations, the earlier results reflect only those who participated in remediation activities; thus, scores of zero for absences during remediation would not be an indicative measure of the overall remediation impact. In particular, some students either did not take all of the quizzes or did not participate in the recitation for remediation purposes; thus, those values were not counted within Table 6. Quiz 2 data were not collectible because of the inability to perform grouping before recitation due to delays in data availability. The number of students not participating in the remediation was lower for initial quizzes and increased during the busy midterm

Table 6. Scores Using Randomly or Intelligently Clustered Groups.

Remediation activity	Measures	Random grouping	Intelligently clustered	Benefit over baseline
Post-Quiz 1 activity	*M*	99%	100%	1.2%
	N	51	54	
Post-Quiz 3 activity	*M*	95.5%	98.0%	2.7%
	N	52	54	
Post-Quiz 4 activity	*M*	79.0%	86.5%	9.3%
	N	49	49	
Post-Quiz 5 activity	*M*	92.5%	94.0%	1.6%
	N	66	26	

Table 7. Chi-Square Distance.

Assessment	Measures	Intelligently clustered	Randomly clustered
Quiz 1	*M*	1.05	0.89
	SD	0.3	0.31
	N	32	23
Quiz 5	*M*	1.03	0.81
	SD	0.22	0.3
	N	32	27

exam period and in the latter part of the semester. The novelty of the quiz/remediation sessions might have inspired students to initially partake in the remediation, while time pressures during the end of the semester could contribute to the 15% variation in student participation. Finally, a latent insight might point to the fact that software engineering topics at the end of the semester tend to be more wordy than mathematical; it might have imposed an artificial ceiling relative to calculation-based content prevalent throughout most other ECE or CS courses.

Chi-square distance analysis. The Auto-PLC grouping technique identified student pairs having high differences in values for achievement assessment of specific skills. As previously described in the Auto-PLC Approach section, to measure widespread differences among learners, the achievement scores listed in the skills matrix measured the chi-square distance between them. This section examines the chi-square values for randomly clustered remediation teams to those for intelligently clustered teams. Table 7 shows the chi-square distances for intelligently clustered peer-learning cohorts from Quiz 1 and Quiz 5, when compared with those that were randomly clustered for the same quizzes. Quiz 1 used pair grouping for its remediation cohorts, while the other quizzes separated into groups of up to four people. The interpretation of the chi-square distance between any pair of cohorts reflects the difference between their mutual skill achievements, whereas the chi-square distance among any four learners leads to a more complex multidimensional representation to convey meaningful interpretations. By roughly 20%, the mean of the chi-square values for the randomly matched samples is less than the mean of the chi-square distance for Auto-PLC's intelligently clustered samples. This affirms the feasibility of grouping students who differ by their mastery of skill levels and illustrates the benefit for the Auto-PLC grouping technique as it reduces instructor workload.

Student Perceptions

To gather student feedback on the course delivery, CBA students engaged in two anonymous, nonmandatory online -approved surveys. At the beginning of the semester, 100 students out of 120 participated in a voluntary presurvey, while at the end of the semester, 40 of the 118 offered students participated in a voluntary postsurvey. The heavy year-end student workloads and the desire of students end the semester attributed to the lower number of postsurvey responses. The Auto-PLC delivered as well as evaluated the completed surveys to analyze student perceptions. The presurvey sought to look at the preexisting student perceptions concerning computer-based quizzes relative to paper-based ones. The postsurvey mainly concentrated on the peer-orchestrated remediation group activities. Most of the survey questions requested responses on a 5-point Likert-type scale: 1= *Strongly Disagree*, 2= *Disagree*, 3= *Neutral*, 4= *Agree*, 5= *Strongly Agree*.

Figure 6 shows student responses to the postsurvey questions. Figure 6(a) addresses student motivation, that is, whether students were more motivated to better prepare for the remediation secessions. More than half, that is, 51% either strongly agreed (13%) or agreed (38%). Because of the grouping of remediation students, some students commented that because they sought to demonstrate their understanding of the course concepts and contribute positively, they purposely prepared for their remediation session. Other results indicated that throughout the course, remediation groups had a positive perception of this type of student engagement. Also, it is interesting to note that there were no negative postquiz responses concerning remediation groups via recitation. It is safe to say that most students appreciated the offered learning intervention.

Figure 6(b) focuses on learning difficult course concepts. The results of the survey indicated that 30% of students responded positively compared with only 6% responding negatively about learning difficult course concepts. This means that the organization and implementation of remediation groups had a positive impact on student perceptions. It is worth pointing out that the students were not required to take the remediation and that the goal of the remediation was to help students better understand concepts related to the questions they missed on the quiz. If they score well in the quiz, many students did not consider coming to the remediation sessions, which would explain why 64% of students responded neutrally about this question.

Figure 6(c) shows student responses as to whether students wished more courses would offer postquiz remediation groups. Overall, the results were quite positive with almost 60% positive responses (38% agreeing and 21% strongly agreeing) by indicating that other courses offer remediation groups. Furthermore, this result showed that postquiz remediation groups generally welcomed as many students sought and needed more assistance. Also, it is important to point out that the response rate of those who were neutral included many students who demonstrated their mastery during the CBAs and did not need additional assistance. Thus, the high level of neutral responses in the survey might indicate that many students wanted to help their peers learn even though they, themselves, did not need it. Last, another promising fact that could

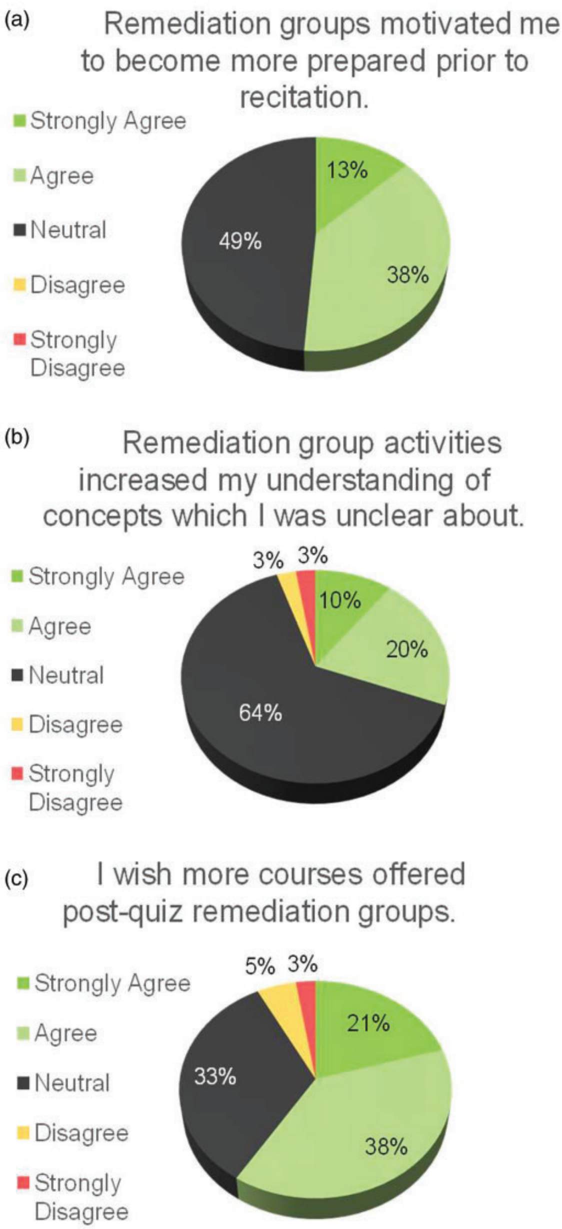

Figure 6. Postsurvey perceptions.

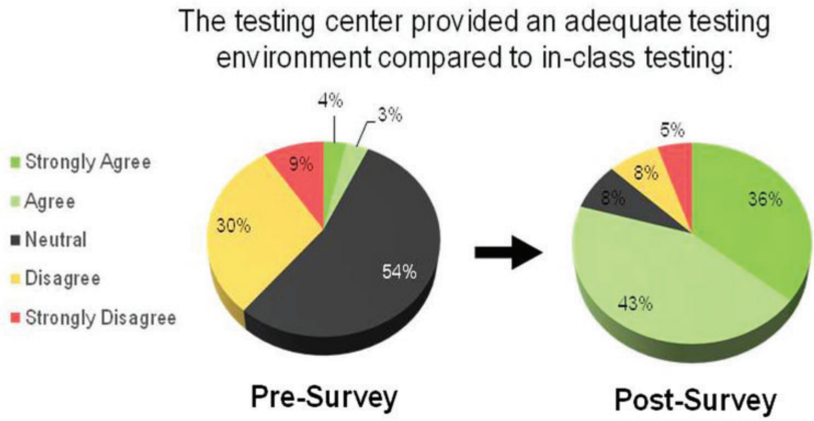

Figure 7. Presurvey to postsurvey shifts in perceptions.

motivate future research is that only 8% of all students had disagreed that more courses should offer postquiz remediation groups, whereas a mean score of nearly 80% was in favor or neutral. This representation is a noteworthy and promising indicator for more implementations of Auto-PLC.

Finally, based on the pre- and postsurveys, Figure 7 shows student attitudes toward the efficacy of the EPC to deliver digitized assessments. The results indicate a favorably shift, whereas at the beginning of the semester, most students were uncertain as to whether it could provide a positive experience compared with the PBA and during the end of semester postsurvey, around 79% agreed including 36% who strongly agreed that it did provide a positive experience. Thus, the use of digitized assessments was both feasible and as an enabler to facilitate ECE and CS learning.

Instructor's Evaluation

Over previous years, the instructor offered the same course in the PBA environment and provided several advantages of the CBA.

Because the quizzes took place in the EPC, the instructor gained an extra 150 minutes of lecture time during the semester because the paper delivery of a midterm would have consumed much time, and reduced grading tasks provided even more time. This allowed extra time for the lecturer to more thoroughly cover difficult subjects as well as introduce current software engineering research topics.

Because the quizzes and remediation recitations occurred in tandem during the same week, it increased attendance during the remediation classes and gave students more time to get to know each other and collaborate face-to-face. Because approximately one half of the software engineering final course grade was project-based, student interaction was especially important and critical in fulfilling this task.

Because the five quizzes and subsequent extra credit remediation quizzes took place during the same week, the CBA, itself, facilitated student learning and provided important benefits. From a course administration point of view, there were less logistical bottlenecks, and most important, it afforded students a second chance to remedy their knowledge gaps. Because the quiz and remediation processes occurred during the same week, it allowed revisiting problems while still fresh in the student's minds. Also, it increased the available GTA exposure time, which was time important for answering students' questions, tutoring, and reviewing EPC computerized tests.

Conclusion

There are fundamental reasons that demonstrate the effectiveness of how remediation groups of two to four students using recent quizzes can advance the learning of ECE and CS subjects. Offered in a remediation session, students who may be struggling to understand a particular set of concepts can substantially benefit. This is especially true for intelligently clustered peer-learning groups, where student clustering occurs using the Auto-PLC to leverage complementary skill efficacies as verified in the various assessments. Although the differences between the randomly and intelligently clustered groups were minor, it would be interesting to explore whether course content that was more mathematically oriented showed greater benefits. Also, to increase the impact of learning, future work could incorporate other distancing metrics, such as *dot product* or *exclusive-OR* for one and zero entries as a means to determine the clustering.

Because the surveys indicated that most of the students were increasingly satisfied with the CBA delivery; instead of groupings based solely on skills, training data of raw quiz, final exam scores, as well as demographic information could allow a machine learner to provide better insights. Even greater insights could result from the analysis of additional data covering several semesters of COP4331 CBA offerings.

The double-blind methodology design employed in this research can be an effective means for assessing the impact of learning interventions without bias. Future work that uses the Auto-PLC will include this delivery approach to a Mechanical Engineering undergraduate course of 230 students on *Heat Transfer Fundamentals* that is awaiting approval. This effort has the vigorous endorsement of the course instructor, the Chair of the Mechanical Engineering Department, and the Dean of the College of Engineering.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Ronald F. DeMara https://orcid.org/0000-0001-6864-7255
Salih S. Bacanli https://orcid.org/0000-0002-2525-5873

References

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Allyn & Bacon.

Chen, B., DeMara, R. F., Salehi, S., & Hartshorne, R. (2017). Elevating learner achievement using formative electronic lab assessments in the engineering laboratory: A viable alternative to weekly lab reports. *IEEE Transactions on Education*, *PP*(99), 1–10. https://doi.org/10.1109/TE.2017.2706667

Craven, J., & Cooper, B. S. (2016). The supervision and mentoring of science teachers: Building capacity in the absence of expertise in classroom science supervisors. *CLEARvoz Journal*, *3*(2), 9–14.

Deibel, K. (2005). Team formation methods for increasing interaction during in-class group work. *SIGCSE Bulletin*, *37*(3), 291–295. https://doi.org/10.1145/1151954.1067525

DeMara, R. F., Chen, B., Hartshorne, R., & Zand, R. (2017a). *Digitizing and remediating engineering assessments: An immersive and transportable faculty development workshop*. Paper presented at the 2017 ASEE Annual Conference & Exposition, Columbus, OH, June.

DeMara, R. F., Khoshavi, N., Pyle, S., Edison, J., Hartshorne, R., Chen, B., & Georgiopoulos, M. (2016, June 26–29). *Redesigning computer engineering gateway courses using a novel remediation hierarchy*. Paper presented at the American Association for Engineering Education National Conference (ASEE-16), New Orleans, LA.

DeMara, R. F., Salehi, S., Chen, B., & Hartshorne, R. (2017b). *GLASS: Group learning at significant scale via WiFi-enabled learner design teams in an ECE flipped classroom. Paper presented at the American Society for* Engineering Education (ASEE) Annual Conference & Exposition, Columbus, OH, June.

DeMara, R. F., Salehi, S., Khoshavi, N., Hartshorne, R., & Chen, B. (2016). *Strengthening STEM laboratory assessment using student-narrative portfolios interwoven with online evaluation*. Paper presented at the American Association for Engineering Education Southeastern Conference (ASEE-SE-16), Auburn, AL, 13-15 March.

DeMara, R. F., Salehi, S., & Muttineni, S. (2016, March 13–15). *Exam preparation through directed video blogging using electronically-mediated realtime classroom interaction*. Paper presented at the American Association for Engineering Education Southeastern Conference, Tuscaloosa, AL.

Epstein, M. L., Lazarus, A. D., Calvano, T. B., & Matthews, K. A. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, *52*(2), 187.

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, *24*(4), 470–497.

Heller, P., & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, *60*(7), 637–644. https://doi.org/10.1119/1.17118

Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, *60*(7), 627–636. https://doi.org/10.1119/1.17117

Henry, T. (2013). *Creating effective student groups: An introduction to groupformation.org*. Paper presented at the Proceeding of the 44th ACM technical symposium on Computer science education, Denver, CO, USA, March.

Jansson, P. M., Ramachandran, R. P., Schmalzel, J. L., & Mandayam, S. (2010). Creating an agile ECE learning environment through engineering clinics. *IEEE Transactions on Education*, *53*(3), 455–462.

Karbalaei, A., Turgut, D., Dagley, M., Vasquez, E., & Cho, H. J. (2018). *Collaborative multidisciplinary engineering design experiences in IoT (Internet of Things) for teachers through summer research site program*. Paper presented at ASME 2018 International Mechanical Engineering Congress and Exposition, November. Pittsburgh, Pennsylvania.https://doi.org/10.1115/IMECE2018-87491

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79–97.

Lin, J.-W., & Lai, Y.-C. (2013). Harnessing collaborative annotations on online formative assessments. *Journal of Educational Technology & Society*, *16*(1), 263–274.

Loughry, M. L., Ohland, M. W., & Woehr, D. J. (2013). Assessing teamwork skills for assurance of learning using CATME team tools. *Journal of Marketing Education*, *36*(1), 5–19. https://doi.org/10.1177/0273475313499023

Meyer, D. (2009). OptAssign-A web-based tool for assigning students to groups. *Computers & Education*, *53*(4), 1104–1119. https://doi.org/10.1016/j.compedu.2009.05.022

Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2004). *Team-based learning: A transformative use of small groups in college teaching* (1st Pbk. ed.). Stylus Publishing.

Muuro, M. E., Oboko, R. O., & Wagacha, P. W. (2016). Evaluation of intelligent grouping based on learners' collaboration competence level in online collaborative learning environment. *The International Review of Research in Open and Distributed Learning*, *17*(2), 40–64.

National Academy of Sciences, National Academy of Engineering, & Institute of Medicine (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. The National Academies Press. https://doi.org/10.17226/11463

Rawson, K. A. (2015). The status of the testing effect for complex materials: Still a winner. *Educational Psychology Review*, *27*(2), 327–331.

Razzaq, L., & Heffernan, N. T. (2006, June). *Scaffolding vs. hints in the Assistment System*. Paper presented at the International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Schurmeier, K. D., Shepler, C. G., Lautenschlager, G. J., & Atwood, C. H. (2011). Using item response theory to identify and address difficult topics in general chemistry. In D. M. Bunce (Ed.), Investigating classroom myths through research on teaching and learning (pp. 137–176). ACS Publications.

Squire, K., & Patterson, N. (2010). *Games and simulations in informal science education* (WCER Working Paper No. 2010-14). http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Srba, I., & Bieliková, M. (2012). *Encouragement of collaborative learning based on dynamic groups*. In A. Ravenscroft, S. Lindstaedt, C.D. Kloos , D. Hernández-Leo (Eds.), 21st Century Learning for 21st Century Skills. EC-TEL 2012. Lecture Notes in Computer Science, vol 7563. Springer.

Srba, I., & Bieliková, M. (2015). Dynamic group formation as an approach to collaborative learning support. *IEEE Transactions on Learning Technologies*, *8*(2), 173–186. https://doi.org/10.1109/TLT.2014.2373374

Teacher Advisory Council. (2009). *Engineering in K-12 education: Understanding the status and improving the prospects*. National Academies Press.

Thomas, A. S., Bonner, S. M., Everson, H. T., & Somers, J. A. (2015). Leveraging the power of peer-led learning: Investigating effects on STEM performance in urban high schools. *Educational Research and Evaluation*, *21*(7–8), 537–557.

Turgut, D., Massi, L., Bacanli, S. S., & Bidoki, N. H. (2017, June). *Multidisciplinary undergraduate research experience in the internet of things: Student outcomes, faculty perceptions, and lessons learned*. Paper presented at 2017 ASEE Annual Conference & Exposition, Columbus, OH.

Vasquez, E., Dagley, M., Karbalaei, A., Cho, H. J., & Turgut, D. (2019). *Collaborative multidisciplinary engineering design experiences for teachers (CoMET) train the trainer model of supports*. Paper presented at ASEE Annual Conference & Exposition, Tampa, FL, June.

Watkins, J., & Mazur, E. (2013). Retaining students in science, technology, engineering, and mathematics (STEM) majors. *Journal of College Science Teaching*, *42*(5), 36–41.

Webb, J. M., Stock, W. A., & McCarthy, M. T. (1994). The effects of feedback timing on learning facts: The role of response confidence. *Contemporary Educational Psychology*, *19*(3), 251–265.

Xiong, X., & Beck, J. B. (2015). *Improving long-term retention level in an environment of personalized expanding intervals*. Paper presented at the Educational Data Mining, June. Madrid,Spain.

Xiong, X., & Beck, J. E. (2014). A study of exploring different schedules of spacing and retrieval interval on mathematics skills in ITS environment. In S. Trausan-Matu, K. E. Boyer, M. Cros, K. Panourgia (Eds.), *Intelligent Tutoring Systems*. ITS 2014.

Lecture Notes in Computer Science, vol 8474. Cham: Springer. https://doi.org/10.1007/978-3-319-07221-0_63

Xiong, X., Wang, Y., & Beck, J. (2015). *Improving students' long-term retention performance: A study on personalized retention schedules*. Paper presented at the Learning Analytics, March. NY: Poughkeepsie.

Zilles, C., Deloatch, R. T., Bailey, J., Khattar, B. B., Fagen-Ulmschneider, W., Heeren, C., Mussulman, D., & West, M. (2015, June). *Computerized testing: A vision and initial experiences*. Paper presented at 2015 ASEE Annual Conference & Exposition,Seattle, Washington, DC, 14 June 2015. https://doi.org/10.18260/p.23726

Author Biographies

Ronald F. DeMara is a professor of Electrical and Computer Engineering at UCF since 1992. He is the UCF Digital Learning Faculty Fellow with research focus on classroom instructional technologies and digitization of STEM assessments. He is the PI of NSF Workshop on Digitally-Mediated Team Learning, Director of UCF's Evaluation and Proficiency Center, and iSTEM Fellow. He completed 300 technical and educational publications, 48 funded projects totaling $12M, and 23 PhD graduates. He received the Effective Practice Award from Online Learning Consortium and the Joseph M. Biedenbach Outstanding Engineering Educator Award from IEEE.

Salih S. Bacanli received his Computer Engineering BS degree from Bilkent University,Turkey. He received his MS in Computer Science degree from University of Central Florida. He is continuing his PhD in Department of Computer Science at UCF. His research interests include opportunistic networks, wireless sensor networks, and co-operative computer science education.

Neda Bidoki received her Computer Engineering BS from Sharif University of Technology, Iran. She received her MS in Computer Science degree from University of Central Florida. She is continuing her PhD in Computer Science Department at UCF. Her research interests include AI and machine learning, the internet of things and co-operative computer science education.

Jun Xu is research scientist at Facebook. He has received his MS and PhD degrees in Computer Science from University of Central Florida (UCF), USA. His research interests include human mobility, agent path planning and machine learning.

Edwin Nassiff is an associate in Research and Data Analytics at the University of Central Florida (UCF) and teaches graduate level courses in Network Science and the Capstone Project in Data Analytics. He is directly involved in as a mentor and teacher in the UCF I-Corps (Innovation Corps), student chapter advisor for ACM, and the Association of DataSci@UCF. His main interests are in emerging technologies, network science, machine and deep learning, entrepreneurship, and software engineering.

Julie Donnelly is a post-doctoral researcher in the Faculty Center for Teaching and Learning at the University of Central Florida (UCF). She earned her PhD in Chemistry from CF, and her research areas include teaching and learning in active learning spaces and chemistry education.

Damla Turgut received her PhD from the Computer Science and Engineering Department of University of Texas at Arlington. Her research interests include wireless ad hoc, sensor, underwater and vehicular networks, cloud computing, smart cities, IoT-enabled healthcare and augmented reality, as well as considerations of privacy in the Internet of Things. She is also interested in applying big data techniques for improving STEM education for women and minorities. Her recent honors and awards include the Women of Distinction Award by the UCF Faculty Excellence Center for Success of Women Faculty in September 2018, the University Excellence Award in Professional Service in April 2017 and being featured in the UCF Women Making History series in March 2015. She was co-recipient of the Best Paper Award at the IEEE ICC 2013. She serves on several editorial boards and program committees of prestigious ACM and IEEE journals and conferences. Her research has been funded through grants from National Science of Foundation, Department of Energy, US Army RDECOM, Florida Center for Cybersecurity, as well as industry grants.