2018 ASEE Annual Conference & Exposition
June 24–27, 2018 Salt Palace Convention Center Salt Lake City, UT
Paper ID #21119

Automated Formation of Peer-learning Cohorts Using Computer-based Assessment Data: A Double-blind Study within a Software Engineering Course

Dr. Ronald F. DeMara P.E., University of Central Florida

Ronald F. DeMara is a Professor of Electrical and Computer Engineering at the University of Central Florida where he has been a faculty member since 1992. His educational research interests focus on classroom and laboratory instructional technology, and the digitization of STEM assessments. He has completed roughly 225 technical and educational publications, 43 funded projects as PI/Co-PI, and established two research laboratories. He serves as the founding Director of the Evaluation and Proficiency Center (EPC) at UCF and is the recipient of UCF's university-level Scholarship of Teaching and Learning Award, Teaching Initiative Program Award, Research Initiative Award, Excellence in Undergraduate Teaching Award, Advisor of the Year, Distinguished Research Lecturer, Marchioli Collective Impact Award, and is an iSTEM Fellow. He received the Joseph M. Bidenbach Outstanding Engineering Educator Award from IEEE in 2008.

Dr. Damla Turgut, University of Central Florida

Damla Turgut is an Associate Professor at the Department of Computer Science at University of Central Florida. She received her BS, MS, and PhD degrees 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, 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 as well as the digitization of STEM assessments. She is PI and Co-PI for NSF-funded REU and RET programs respectively. She co-led iSTEM Fellows program at UCF during 2016-2017 AY. Her recent honors and awards include 2017 University Excellence in Professional Service Award, 2017 Teaching Incentive Program Award 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. Dr. Turgut serves as a member of the editorial board and of the technical program committee of ACM and IEEE journals and international conferences. She is a member of IEEE, ACM, and the Upsilon Pi Epsilon honorary society.

Dr. Edwin Nassiff, University of Central Florida

Edwin Nassiff is an Adjunct Professor in the Computer Science department at the University of Central Florida (UCF) where he has been an Adjunct since 2012. He has taught both graduate and undergraduate level courses including software engineering, cloud computing, security, and data base management. He has held senior level positions as a CIO, CTO, and Technical Director with the US Government and Lockheed Martin Corporation. He has mentored for the UCF I-Corps, an organization assisting scientific researchers transform their research ideas into commercial entities. His research interests include enterprise architecture, data base modeling, computer security, and software and systems engineering.

Salih Safa Bacanli, University of Central Florida

Salih Safa Bacanli is PhD student at Department of Computer Science, University of Central Florida (UCF). He received his MS degree in Computer Science from UCF and BS degree in Computer Engineering from Bilkent University, Turkey. His research interests include computer engineering education, opportunistic networking routing, wireless sensor network routing and security. He is member of Upsilon Pi Epsilon honorary society,ASEE and Order or Engineer.

Neda Hajiakhoond Bidoki, University of Central Florida

Neda Hajiakhoond Bidoki is a Ph.D student at the Department of Computer Science at University of Central Florida. Her research interests includes wireless network networks, mobility models, machine learning, data analysis as well as considerations of privacy in the Internet of Things. She received her M.Sc. in Network Engineering from Amirkabir University of Technology and her B.Sc. in Information Technology from Sharif University of Technology.

Mr. Jun Xu, University of Central Florida

Jun Xu is pursuing the PhD degree in Computer Science from the Department of Computer Science at the University of Central Florida. He received his MS degree in Electrical Engineering from Beijing University of Posts and Telecommunications, China. His research interests include mobility models, agent path planning, and machine learning techniques applied to both large-scale autonomous and learning analytics systems.

Automated Formation of Peer Learning Cohorts using Computer-Based Assessment Data: A Double-Blind Study within a Software Engineering Course

Ronald F. DeMara[1], Damla Turgut[2], Edwin Nassiff[2], Safa Bacanli[2], Neda H. Bidoki[2] and Jun Xu[2]
{[1]Department of Electrical and Computer Engineering, [2]Department of Computer Science}
University of Central Florida, Orlando, FL 32816-2362

Abstract

An approach is developed to integrate the complementary benefits of digitized assessments and peer learning. The research hypothesis is that each student's assessment data at the fine-grained resolution of correct/incorrect question choice selections can be utilized to partition learners into effective peer learning cohorts. A low overhead approach is explored along with its associated tool, referred to as *Automated Peer Learning Cohorts (Auto-PLC)*. The objective of Auto-PLC is to increase scalability to deliver peer-based learning. First, digitized formative assessments are delivered in a computer-based testing center. This enables automated grading, which frees-up the instructor's and teaching assistants' workloads to become reallocated to recitation sessions for higher-gain learning activities, such as peer-based remediation sessions. Second, within the recitations held following each formative quiz, students are afforded an opportunity to complete a remedial assignment. Auto-graded results of formative assessment submissions undergo Auto-PLC's statistical clustering routines using Excel macros and Python scripts to partition the class into four-person peer learning cohorts having mutually-complementary knowledge gaps and skill efficacies. Within each peer learning cohort, students solve together an assigned remedial problem during the recitation session. Thus, students who have already acquired a particular skill become paired together with students who are still acquiring that same skill, and vice versa. This also aids scalability to large enrollments within Electrical and Computer Engineering (ECE) and Computer Science (CS) courses by maximizing opportunities for students to teach each other the material which they still need to learn.

The motivation, design, and outcomes for Auto-PLC are presented within the required undergraduate course *COP4331: Processes for Object-Oriented Software Development* at a large state university. To evaluate effectiveness, a double-blind IRB-approved study has been conducted in COP4331 involving 206 students. All enrolled participated identically, except for their assignment to either randomly-formed or intelligently-clustered remediation groups. At the end of the semester, all students completed an identical Final Exam to provide a basis by which to compare their relative achievements. The data collected expounds upon the details of Auto-PLC's impact towards achievement on a topic-specific basis. Additionally, learners' perceptions of digitized assessments and participation in recitation-based peer learning cohorts are discussed.

1.0 Motivation

Throughout the industrial practice and instructional roles across engineering and computer science disciplines, the activities of team design, group problem solving, and project collaboration have always been a prominent and defining attribute of STEM fields. Especially in the last two decades and into the foreseeable future, team design skills are receiving increasing importance as complexity of science and engineering marches ever forward [3]. The rising tide of complexity necessitates future graduates at all levels within STEM fields to function effectively as disciplinary specialists who work together closely and frequently during most phases of product development and research. While always an integral element of STEM curricula, the need and benefit for learners to become immersed in collaborative learning activities have become highlighted in order to elevate needed proficiencies in team-based skills [5]. Thus, the priority for advancing forward-looking educational technologies that demonstrate the most significant potential to advance team-based instruction is vital and broadly impacts STEM fields where lab partners, group project teams, collaborative design projects, and even Senior Design courses rely heavily on team-based learning.

The research herein explores the extension of peer learning activities by leveraging digitized formative assessments within certain ECE and CS courses. A promising pathway can be to focus first on optimizing the formation of learner teams based on data mining or machine learning algorithms which may be tuned to reach diverse learners by forming teams having complementary skills. These may include at-risk learners that can benefit from placement on teams with others who have already demonstrated that skill. This paper investigates the utilization of a combination of promising technologies to advance digitally-mediated team learning, starting with learning analytics to form more effective learning teams. It explores the development of cyber-assisted peer learning approaches for interaction and remediation of student teams via focused educational data mining of the already collected and available formative assessment data. Thus, a low-overhead approach to increase learning becomes feasible in such settings.

2.0 Research Objectives and Approach

The feasibility of digitized assessments within Engineering and Computer Science disciplines continues to receive increasing attention during recent years [6] [7] [8]. Challenges facing digitization of assessments within technical curricula include equitable mechanisms for partial credit, scalable submission and grading of handwritten work, and evaluation of creative design aspects within the constraints of contemporary Learning Management Systems (LMSs) [9]. Herein, lockdown proctored computer-based testing was evaluated as an enabling instructional technology to reallocate low-gain grading tasks of the instructor and Graduate Teaching Assistants (GTAs) towards conducting more impactful recitation sessions by mentoring remedial exercises with *purposely-formed peer-learning cohorts*.

To address this objective, both conventional in-class paper-based exams as well as computer-based exams were delivered. Computer-based testing occurred within an *Evaluation and Proficiency Center (EPC)* [9], which is a College of Engineering and Computer Science specific testing and tutoring center. The EPC resides in a once-vacated open computing lab and is staffed in-part by GTAs who became freed-up due to their abridged grading workloads. EPC-based delivery allows students to complete exams asynchronously in a secure manner at a time convenient to the student and with little burden on the part of faculty, as detailed in Section 2.

Skill-optimized peer-enhanced learning cohorts were formed having complementary knowledge gaps and skill efficacies. The learner cohorts were constructed automatically via Chi-Square test clustering analysis using the formative assessment results which have been accumulated to-date in the course. Within each peer-enhanced learning cohort, students who have already acquired a particular skill become matched together with those students who are still acquiring that same skill, and vice versa. This aids scalability to large enrollments by maximizing opportunities for students to teach each other material which they need to learn. Students were allocated a two-day window early in the week to schedule and take their computer-based quiz asynchronously in the EPC. EPC testing uses question ‘clones’ to help prevent crosstalk between test delivery instances. On Friday of the same week, the students were clustered (randomly or intelligently) to conduct various remediation activities for extra credit in groups ranging from two to four students. The intelligently-clustered remediation groups were formed based on their individual quiz results where students who missed a portion of the quiz were matched with students who answered that portion correctly and vice-versa.

Specifically, the use of computer-based testing via proctored lockdown delivery within a Computer Science and Engineering undergraduate core course is studied herein using the double-blinded study design shown in Figure 2.1. First, digitized assessments were integrated into the undergraduate course entitled *COP4331: Processes for Object-Oriented Software Development*, at a large enrollment state university. During the Fall 2017 semester under IRB approval, a crossover study randomly-partitioned COP4331’s students into one control group and two intervention groups. The lecture and laboratory components were conducted identically for all cohorts. Within Figure 2.1, it is depicted that the cohorts’ formative assessments were delivered either in the EPC testing center or via paper-based assessments. Within each cohort, the interventions of computerized delivery, as well as random or intelligently-clustered peer learning groups, continued during successive topic modules on a mutually-exclusive basis. An identical paper-based Final Exam was delivered to all students. The data collected included various pre/post surveys of students’ perceptions, student achievement as scores on a range of formative and summative assessment, and time logs of the instructor and GTAs. These were analyzed to investigate the topic-specific effects of digitized assessment as described in Section 5. Using the aforementioned study configuration, the following Research Question was targeted:

Do peer learning cohorts which are formed via assessment-driven clustering outperform peer learning cohorts which are formed randomly?

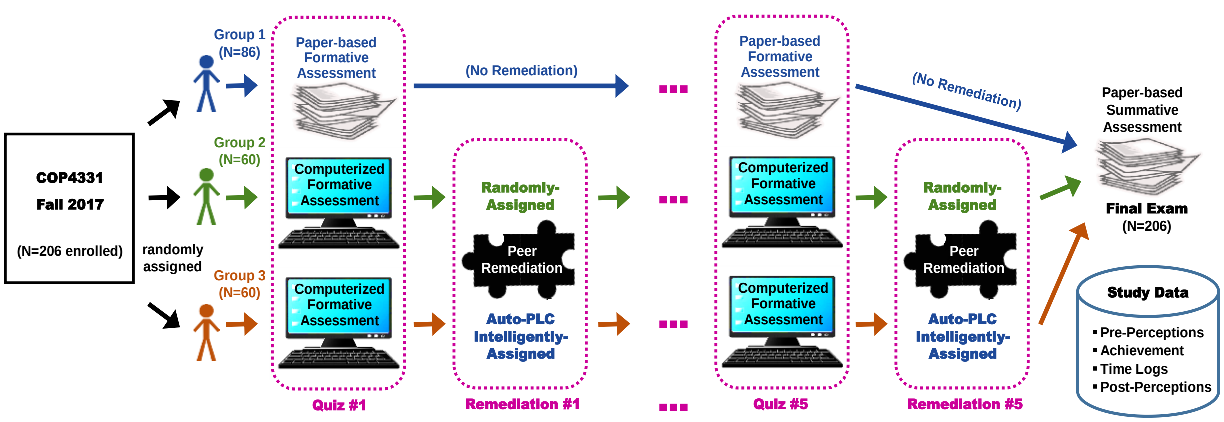

Figure 2.1: Double-Blind Study Design. Participants are randomized to either *Paper-based* or *Computerized* testing for each of 5 *Formative Assessments*. Peer learning cohorts are offered to Groups 2 & 3 assigned either randomly or using Auto-PLC, respectively. All participants complete an identical *Summative Assessment*.

To investigate the Research Question, we analyzed the mean scores of post-remediation assessment results. Since the recitation sessions contain both randomly-grouped and intelligently-clustered learning teams as shown in Figure 2.1, there were opportunities to measure the differences in achievement using random matching to form remediation cohorts as compared to Auto-PLC intelligently-clustered peer learning cohorts. Analysis was performed by a study assistant, thus remains blind to students and the professor.

Figure 3.1: Digitized assessments in 120-seat EPC.

3.0 Selected Related Works

3.1 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 within Engineering and Computer Science curricula. Referred to as an *Evaluation and Proficiency Center (EPC)* [9], this infrastructure promulgates an integrated testing and tutoring methodology, which has been used to support a broad range of STEM programs [10]. Digitization enables auto-grading of assessments, which frees up graders for tutoring to allow high-gain teaching and learning activities. The 120-seat capacity EPC supports assessment and enhanced remediation. Low cost testing can also enable the increased use of formative assessments to provide rapid feedback and an effective approach for mastery by students and guiding instructors [11]. Moreover, the *Testing Effect* of 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 [12]. Thus, proctored testing can invoke the Testing Effect more frequently through in-person digitized formative assessments, in lieu of low-gain homework submissions. This leads the learner to elevate their engagement and to increase the ownership of their learning outcomes. Figure 3.1 shows learners taking digitized formative quizzes delivered at their preferred appointment times. 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 [13]. Services provided by the EPC infrastructure include quiz appointment scheduling, student authentication, stowage of unauthorized materials, pre/post-tutoring, and self-paced solution review. Faculty services provided by the EPC include turnkey delivery of the primary and make-up exams, proctoring, scratch paper scanning, auto-grading, gradebook entry, video/attendance recording, and others.

Figure 3.2: EPC Peer Learning Cohort.

3.2 Formation of Peer Learning Cohorts as a Remediation Strategy

Beyond the delivery of tests, additional Instructor and GTA-guided remediation also becomes possible when adopting digitized assessments in ECE and CS courses. This can offer learning advantages versus paper-based testing which typically incurs a week turnaround delay for the return of graded submissions. It can also pioneer a novel *Peer Learning Cohort* technique, which self-motivates learners in a quest for partial credit to explain the problem-solving flow in their formative assessment submissions. This is accomplished within the EPC by scanning in any student hand-written scratch worksheets composed during assessment. Later, these scanned sheets can be used by the student and GTA tutors together while the student explains their solution in their own words. This first-line remediation between the GTA tutors and students can result in a regrading of the assessment and provides additional feedback to the instructor. Thus, computerization of assessments can increase student engagement through in-person tutoring interwoven with assessment via Socratic discussions which foster metacognition. Figure 3.2 depicts learners conducting secure self-paced review of their formative quizzes by engaging in Socratic questioning to gain partial credit based on scanned scratch sheets. Thus, the EPC extends the promising aspects of an "Open Tutoring Center" with tutors available for targeted assistance [14] where the authors pointed out the absence of an effective, integrated, and verifiable assessment methodology. Furthermore, Auto-PLC builds on the benefits of peer instruction [15] to extend the advantages of previous works [16] using learning analytics to form the peer learning cohorts, as identified below.

Currently, many pedagogical reforms in STEM provide enhanced learning processes and environments, but often do not consider learner team performance, which limits their effectiveness. This project investigated the potential of a broad vision leveraging peer learning, realized through learner modeling and data mining techniques, to simultaneously increase engagement, quality, creativity, and integrity, beyond existing pedagogies [17] [18] [19]. This occurs during recitation by the responsiveness to learner behaviors thus supporting interactions

Table 3.1: Selected related works on intelligent grouping of learner teams.

Approach and/or Tool	**Application Area**	**Group Input Data**	**Methods Employed**	**Group Setting**	**Outcomes**
Dynamic formation based on Group Technology [1]	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 [2]	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 [4]	Computer science education	Prior student knowledge, Personality types	Encouraging Interaction	Short-term in-class groups	Collaborative learning quality improves. Students' attitudes towards in-class group work increase 30%
Auto-PLC *(developed herein)*	Software Engineering/ Programming	Formative Assessments	Chi-Square Categorization	Recitation Peer Learning	1.2% to 9.3% achievement increase post-remediation

with instructional processes and events. Through the leveraging of student achievement data, cyber-enabled adaptive team composition, and real-time monitoring to sustain instantaneous modeling of the learner, it is likely to realize 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 [20-22], Heffernan [23, 24], Koedinger [25], Salame [26] and others identify tradeoffs in learning outcomes with online formative assessments through immediate feedback, which is useful for allowing for reflection whereby the student use of feedback becomes a tool for continuous growth [27-29].

Related works utilizing dynamically-formed peer cohorts are summarized in Table 3.1. For instance, Srba et al. investigated the creation of dynamic short-term team formations irrespective of student achievement [2]. This approach considered students' 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 the proposed method achieved a higher collaboration quality in comparison with the reference groups. Henry proposed a novel group formation system: **`groupformation.org`** [2]. This system provides an interactive environment which allows the instructor to experiment with different grouping parameters and algorithms. Informal surveys and discussions with students were used to evaluate the success of this system. In their experiment, most groups worked well together, successfully completed a semester project, and reported high satisfaction with their group learning activities.

Deibel proposed two team formation methods for in-class group work for CS courses [4]. His goal was to increase interaction and promote participation amongst students. The first method, called the latent jigsaw method, assigns groups based on prior student knowledge. The second method uses Felder-Silverman learning styles to promote participation by considering students personality types. Both methods performed very well in terms of quality of collaborative learning and students' attitudes towards in-class group work. Overall, there appears to be significant potential for learner team formation systems to be advanced further along the perspectives of learning benefit and automation. Both of these objectives are addressed herein.

4.0 Auto-PLC Approach

4.1 Methods Employed

Herein, the Testing Effect [30] is leveraged herein for collaborative learning. Incorporating collaborative learning activities (e.g. Think-Pair-Share, peer tutoring, and "jigsawed" learning) can contribute to learners' content knowledge while cultivating vital 21st Century cross-cutting skills required for long-term success [31]. Past research recognized that grouping strategies can elevate academic achievement in learning groups [32]. Herein, groupings remain fluid throughout the course as indicated by learners' needs, which is a distinctly novel approach. While there has been some success with computer-assisted groupings, groupings which are static can yield mixed results [33]. We started our matching process by isolating the quiz scores of each student for all questions and illustrate an example case in Quiz 1 within the COP4331 course. Quiz 1 contained seven questions in true/false, multiple choice, and matching formats. We created a Questions

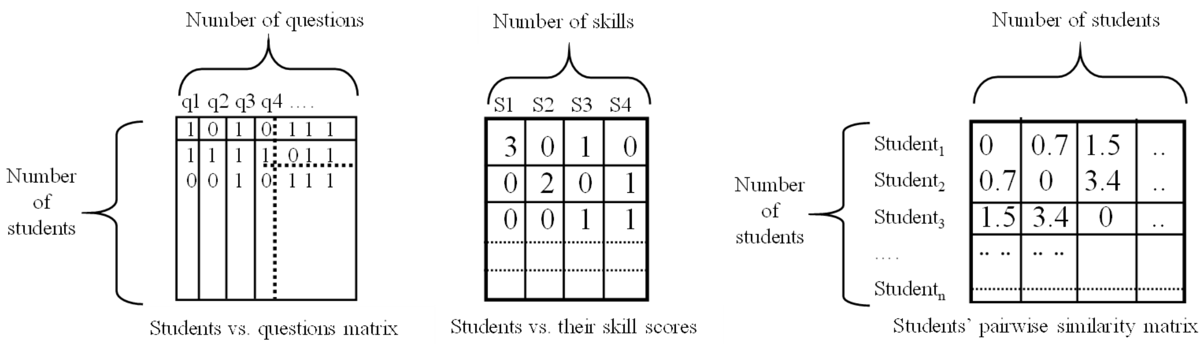

Figure 4.1: Intelligent clustering of students to form peer learning cohorts using their detailed formative quiz results.

Matrix elaborating which questions were answered correctly by each student. That matrix had as many columns as questions, i.e. seven in this example, and had as many rows as students, i.e. 116 in this example. The questions matrix contained only one or zero entries. Seven questions were mapped onto four skills which were: program-level testing (three questions), integration testing (two questions), programming practice (one question), and box testing (one question). A skill matrix of students was created which contained as many columns as skills and as many rows as students. The maximum value of the entry of each cell is based on the skill set. Programming level testing cells could have at most three, as there were only three questions utilizing this skill, whereas programming practice cells could have at most one. Using the skill matrix scores of the students, the Chi-Square method was used to compare students where the Chi-Square distance gave a number to assess the similarity of two students. If the Chi-Square distance was equal to zero, it implied that the students' skill matrices are identical. As distance grows, the similarity between them decreases according to:

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

As depicted in Figure 4.1, Auto-PLC constructs a square matrix that has as many columns and rows as number of students. Each row represents a student. For a specific row the values on the columns shows the Chi-Square distance of that student with all other students. Auto-PLC selected a random student (row) and obtained the student (column) who has the highest Chi-Square score. Thus, Auto-PLC selected the student whose skill score was the most distant. The farthest skill score student was iteratively put through the same process until group of four was formed. Once Auto-PLC found all members of a group, it removed all of the rows and columns of the members so that it could start creating another group using the same algorithm. For Quiz 4, Auto-PLC created groups of four learners, whereas for Quiz 5 student pairs (groups of two)

Table 4.1: Skills assessed in quizzes used to construct the skill matrices defined in Figure 4.2.

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 & Cohesion, • System-level Testing	• Programming Principles • Integration Testing • Box Testing • Program-level Testing	• System-level Testing, • Testing Team • Simulator • Testing and training documents

were formed so that we could see what effects group size had. For student pairing, Auto-PLC simply identified students who differed the most in the Chi-Square distance and filled a pair with that criteria and then removed the pair of students from the matrix. The purpose of picking the most diverse students was to attain a preferred grouping of students who achieved high scores for each skill with those who achieved lower score for each of the assessed skills. The skills matrix construction process was conducted for each assessment in order to form intelligent clustering, see Table 4.1.

4.2 Digitized Assessment and Remediation Design

In order to investigate the effectiveness of Computer-Based Assessment (CBA) relative to Paper-Based Assessment (PBA), one delivery format was utilized for one section and the other for the other section. Results were compared to determine the difference between the two sections. PBAs were delivered in a traditional classroom setting with teaching assistants serving as in-class proctors. Computer-based exams were delivered in the designated testing center called the Evaluation and Proficiency Center (EPC), where proctoring was provided by the EPC staff. In the EPC, Internet access was restricted, and a lockdown browser was used to ensure test integrity and security.

The EPC quiz questions were cloned to help prevent crosstalk among students over the two-day examination period. In addition, the Canvas-based Learning Management Systems (LMS) was used to deliver the quiz. Five quizzes were scheduled throughout the semester. For each quiz, of the 206 students, approximately 42 percent of the students underwent PBA while the other 58 percent were under the CBA. Paper-based and computer-based quizzes were comprised of the same questions with identical content and format. The quizzes consisted of five to ten questions, with a time limit of 30 minutes. Three forms of questions were used including multiple choice (single and multiple answers), matching, and true and false.

4.3 Scoring Procedures and Implications

There were five quizzes given to the students throughout the semester, which eliminated two midterm exams given in previous semesters. As mentioned previously, the CBA quizzes had to be taken on either of two consecutive days at the beginning of the week followed by a remedial quiz given on Fridays of the same week wherein students were given extra credit for the correct answers on the remediation activity. Several concerns about this format arose. One 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?* Although there was no formal study of this issue, we observed that many students who performed well were eager to teach their mastery to those who did not. The fact that the remediation was extra credit and not graded as a quiz led to an environment of low or, no risk, and probably lowered the barriers of reluctance. Students who scored better gained leadership skills, reinforced their own mastery, and felt some sense of accomplishment via contributing to the success of a fellow student. In addition, the higher scoring students were probably motivated by the fact that since their extra credit score was coupled with a lower scoring one, a positive remediation extra credit score could make a difference between a plus or minus on their overall semester grade. In absolute terms, it also elevated their soft-skills for job interviews, presentations, etc. as survey results indicate in

Section 5. For the PBA, the quizzes included various questions for which some students could receive partial credit since the graders graded them manually whereas the majority of the questions from paper-based exams were converted to digital format for CBA students. Therefore, the scoring on the CBA was not given partial credit which could have had an effect on the comparisons between the PBA and CBA. Computer-based exams were graded by a hybrid machine and human-based approach. The CBA students' used scratch paper while taking their exams that were scanned and saved, which could be used later during a score clarification process with a GTA. Grading was initially fulfilled automatically by the LMS, but students could visit their GTA to review their exams coupled with their scratch paper, and if applicable, could obtain partial credit. This capability was possible because of extended GTA office hours which was made possible due to their reduced grading workloads.

5.0 Delivery Results

The results of the study are presented to address the Research Question identified in Section 2.0. To assess effectiveness, a double-blind IRB-approved study had been conducted in COP4331 with 206 undergraduate students during the Fall 2017 semester. Of that total population, 86 students comprised the control group which underwent PBA without remediation, which has been depicted earlier as Group 1 within Figure 2.1. The remaining 120 students enrolled in this senior-level course underwent CBA. This was split into Group 2 having randomly-formed remediation matching and Group 3 having intelligently-clustered remediation matching. All who were enrolled participated identically, except for these differences. At the end of the semester, all students completed an identical final paper-based exam to provide a basis by which to compare their relative achievements.

5.1 Data Collection and Analysis

Both quantitative and qualitative data were collected. Quantitative data was collected through students' assessment scores, self-reported survey results, and the GTAs' time log. Qualitative data was collected through the instructor's reflection. The qualitative data was used to interpret and provide support for the results from the quantitative analyses [34].

With students' permission, scores of students from both EPC and in-class settings on all five quizzes, together with a Final Exam were de-identified. Student scores from both Control and EPC cohorts, including data on free response questions from the quizzes were compared and analyzed. This was used to investigate whether the students gained learning benefits by participating in peer-remediation groups during recitation sessions. To investigate a meaningful effect of remediation on learning outcomes, the quiz submission scores of learners who achieved within the lower quartile were examined. Table 5.1 indicates the results of achievement on the Final Exam for students who

Table 5.1: Final Exam Achievement with and without Remediation.

Post-Quiz 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%
Non-Recipients	77.5%	73.1%	75.0%	75.2%	72.3%
Difference	1.7%	5.2%	2.5%	1.6%	9.0%

received remediation compared to those who did not receive remediation. Their Final Exam scores are listed as the lowest 25% of scores on Quiz 1, Quiz 2, and up to Quiz 5, respectively, within Table 5.1. Specifically, the difference was computed between those who received the corresponding remediation following each quiz, as compared to those who did not receive the remediation corresponding to that quiz, using the course Final Exam as a uniform summative instrument. Results indicate an increase in Final Exam score from about 1.7% for remediation following Quizzes 1 and 4 ranging up to 9.0% for remediation following Quiz 5, which is significant.

The Research Question posed within Section 2.0 postulated whether peer learning cohorts which are formed intelligently via assessment-driven clustering, outperform peer learning cohorts which are formed randomly. The scores listed in Table 5.2 for the formative assessments indicate that the intelligently-clustered students' groups achieved higher scores by participating in peer-remediation groups during recitation sessions, in comparison to randomly-clustered groups within all quizzes. Thus, the preliminary data gathered within this study also supports the positive response to the Research Question within that context, albeit modest which we believe is at least in-part due to the reasons mentioned therein. Expanding further, whereas each student was randomly-clustered and intelligently-clustered roughly equally, comparable achievement results at an aggregate-level of granularity can be an expected result for the Research Question, as seen in the Final Exam results listed in Table 5.1. A further discussion of the reasons as to why these transpired, will be provided within the Conclusion Section of this paper.

Out of the 120 students comprising both intervention cohorts in total, these results only reflect those who participated in remediation activities, whereas scores of zero for absences during remediation would not be an indicative measure of remediation impact. In particular, a few students either did not take the quiz and/or did not participate in the recitation for remediation, thus those values were not counted within Table 5.2. Quiz 2 data was not collectable due the inability to perform grouping prior to recitation because of data availability delays. The number of students not participating in the remediation was lower for initial quizzes and increased during the busy midterm exam period later on in the semester. The novelty of the quiz/remediation sessions might have inspired students to take the remediation at the beginning of the semester while time pressures during the end of semester coupled with the students' other course loads, could contribute to +/-15% variation in participation. Additionally, the average quiz scores varied weekly whereas a range of group activities were afforded as the semester progressed through various technical topics. Finally, a latent insight suspected by the authors is that the subjective topic content in software engineering may impose a ceiling on impact relative to calculation-based content more prevalent in other engineering courses.

Table 5.2: Scores using randomly or intelligently-clustered groups.

Remediation Activity	**Measures**	**Random Grouping**	**Intelligently-Clustered**	**Benefit over Baseline**
Post-Quiz 1 Activity	Mean N	99% 51	100% 54	**1.2%**
Post-Quiz 3 Activity	Mean N	95.5% 52	98.0% 54	**2.7%**
Post-Quiz 4 Activity	Mean N	79.0% 49	86.5% 49	**9.3%**
Post-Quiz 5 Activity	Mean N	92.5% 66	94.0% 26	**1.6%**

Chi-Square Distance Analysis: The Auto-PLC grouping technique was piloted by identifying student pairs having high difference values in achievement on specific skills assessed in the formative assessments. To measure the difference between learners' achievement according to the entries within the skills matrix, the Chi-Square distance metric was utilized as previously defined in Section 4.0. In this subsection, the selectivity of such metric is examined by comparing Chi-Square values of randomly-clustered remediation teams to those formed automatically by Auto-PLC. Table 5.3 shows the Chi-Square distances for intelligently-clustered peer learning cohorts from Quiz 1 and Quiz 5, as compared to those which were randomly-clustered. The results of these quizzes are listed where the remediation quiz groups consisted of two learners, while the other quizzes piloted groups of up to four people. Whereas the Chi-Square distance between a pair of people can be interpreted as the difference between their mutual skill achievements, the Chi-Square distance between four learners leads to a more complex multi-dimensional representation to convey a meaningful interpretation. As indicated, 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 intelligently-clustered samples by roughly 20%. This illustrates the benefit for the Auto-PLC grouping technique. It also affirms the feasibility of grouping the students who differ from each other in terms of their demonstrated skills on the computer-based formative assessments without any additional instructor workload.

Table 5.3: Chi-Square Distance.

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

5.2 Student Perceptions

Student Pre- and Post-Surveys were administered and analyzed regarding Auto-PLC perceptions of the participant learners. Two anonymous IRB-approved online surveys were released to all students, i.e. at the beginning (N=100 of 120 responding) and at the end of the semester (N=40 of 118 responding), to gather student feedback on test delivery. The former sought the preexisting perceptions of students on computer-based exams relative to paper-based exams.
An anonymous post-survey was administered to gather student perceptions. 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 5.1 shows the student self-reported responses to the post survey questions. As shown in Figure 5.1 (a), in terms of whether students were more motivated to become more prepared prior to

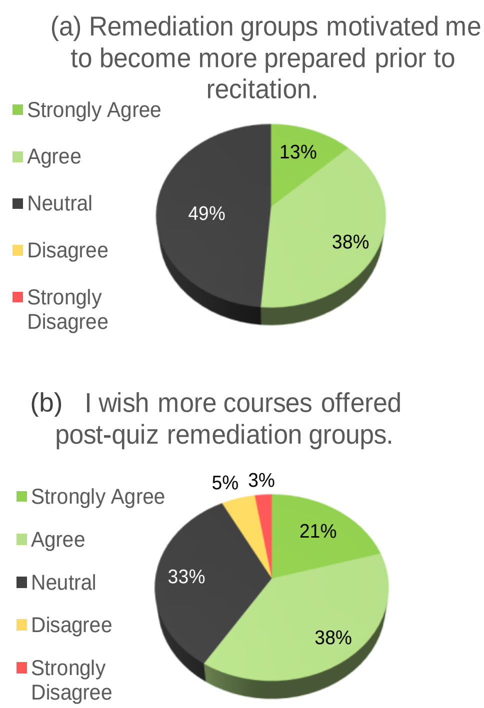

Figure 5.1: Post-survey perceptions.

remediation, more than half, i.e. 51% either strongly agreed (13%) or agreed (38%). Since each student was paired with one or more other students, most students wanted to show they had a good understanding on the course concepts and thus contribute positively to their remediation groups. This result indicated that throughout this course remediation groups contributed positively to student engagement. In addition, it is interesting to note that there were no negative responses. In other words, all students felt the potential for post-quiz remediation groups via recitation as a motivating factor.

Figure 5.1 (b) shows students' responses to whether they wished more courses would offer post-quiz remediation groups. Overall, the results are quite positive with almost 60% of students, i.e. 21% strongly agree and 38% agree, indicating that their interest in having other courses offer remediation groups after each quiz. This result further indicated that the post-quiz remediation groups developed herein were both welcomed and sought after amongst students who needed more assistance. Such remedial activities can help students learn difficult materials and potentially retain at-risk learners. It is important to point out that the response rate of those who were neutral included those other students who had already demonstrated their mastery of the concepts during the computer-based assessment and may have not seen the need for remediation. Thus, a neutral perception of responses from the survey is commensurate with a level-headed or even inspiring outlook of the overall process even from those who helped their peers learn even though they, themselves, did not need assistance. That fact itself is also promising and motivates future extensions. Specifically, only 8% (3% plus 5%) of all enrolled students had disagreed that more courses should offer post-quiz remediation groups, whereas a mean score of those in favor or neutral was around 80%. This representation is a worthy and promising indicator for the overall realization of Auto-PLC and its delivery within the setting described.

6.0 Conclusion

Remediation groups of two to four students after quizzes can advance learning of certain ECE/CS topics and be received favorably. Remediation quizzes were embraced favorably by students who were struggling to understand a particular set of concepts. In addition, the remediation approach appears to be effective and well-liked for this content by both accomplished and struggling students. Using intelligently matched groups, students clustered by Auto-PLC can increase the potential to leverage complementary skill efficacies as demonstrated in the digitized assessments and increases the likelihood to advance each other through collaborative learning. Although the differences between randomly and intelligently-clustered groups were modest for the primarily verbal-dominant COP4331 questions/quizzes, it would be interesting to explore whether more analytically-based content further magnifies these benefits. Also, to increase the impact of grouping, future work could explore different distance metrics, such as dot product or exclusive-OR for one and zero entries. Since the surveys indicated that most of the students were increasingly satisfied with CBA for selected ECE content; instead of grouping based on skills, raw question scores could be derived via deep learning methods to leverage further insights. Finally, since Auto-PLC intelligent-driven clustering remains transparent to the instructor using machine generated team lists for cohorts irrespective of pairing strategy, the double-blind study design used in this research can be an effective means for assessing the impact of learning interventions without bias.

Beyond an assessment-level granularity analysis done for this preliminary conference paper study, a further detailed analysis can be undertaken to discern skill-level differences via the summative assessment results using the Final Exam scores themselves as a lumped value result. Thus, as future work beyond the page limitations of this preliminary conference paper, we will breakdown the matching of the skillsets in the summative assessment to compare outcomes at a finer granularity which will allow us to ascertain the details of efficacy across a palette of learning outcomes. We are also extending Auto-PLC delivery to a Mechanical Engineering undergraduate course on *Heat Transfer Fundamentals* where IRB documents have been submitted and its delivery is planned during the Fall 2018 semester to 230 students who will collectively comprise the control and intervention groups. This effort has the vigorous endorsement of the Course Instructor, the Mechanical Engineering Department Chair, and the College Dean.

References

1. I. Srba and M. Bieliková, "Encouragement of Collaborative Learning Based on Dynamic Groups," in *Proceedings of European Conference on Technology Enhanced Learning*, September 18 - 21, 2012, Saarbrücken, Germany.
2. T. R. Henry, "Creating effective student groups: an introduction to groupformation.org," in *Proceedings of the 44th ACM Technical Symposium on Computer Science Education*, March 6 - 9, 2013, Denver, Colorado.
3. J.-W. Lin and Y.-C. Lai, "Harnessing collaborative annotations on online formative assessments," *Journal of Educational Technology & Society*, 2013. 16(1): p. 263-274.
4. K. Deibel, "Team formation methods for increasing interaction during in-class group work," *SIGCSE Bull*., 2005. 37(3): p. 291-295.
5. R. F. DeMara, S. Salehi, B. Chen, and R. Hartshorne, "GLASS: Group Learning At Significant Scale via WiFi-Enabled Learner Design Teams in an ECE Flipped Classroom," in *Proceedings of American Society for Engineering Education Annual Conference & Exposition*. June 25 – 28, 2017, Columbus, OH.
6. C. Zilles, J. Bailey, B. B. Khattar, W. Fagen, C. Heeren, and M.W. David Mussulman, "Computerized Testing: A Vision and Initial Experiences," in Proceedings of American Society for *Engineering Education Annual Conference & Exposition*. June 14 – 15, 2015, Seattle, WA.
7. K. D. Schurmeier, 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.
8. B. Chen, 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*, Feb, 2018. 61(1): p. 1-10.
9. R. F. DeMara, 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 Conference & Exposition*, June 26 – 29, 2016, New Orleans, LA.
10. R. F. DeMara, B. Chen, R. Hartshorne, and R. Zand, "Digitizing and Remediating Engineering Assessments: An Immersive and Transportable Faculty Development Workshop," in *Proceedings of American Association for Engineering Education National Conference & Exposition*, June 2017, June 25 – 28, 2017, Columbus, OH.

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

12. K. A. Rawson, “The status of the testing effect for complex materials: still a winner,” *Educational Psychology Review*, 2015. 27(2): p. 327-331.

13. R. F. DeMara, S. Salehi, and S. Muttineni, “Exam Preparation through Directed Video Blogging using Electronically-Mediated Realtime Classroom Interaction,” in *Proceedings of American Association for Engineering Education Southeastern Conference*. March 13 – 15, 2016, Tuscaloosa, AL.

14. National Academy of Science, Rising above the gathering storm: Energizing and employing America for a brighter economic future. 2007, National Academies Press Washington, DC.

15. J. Watkins and E. Mazur, “Retaining students in science, technology, engineering, and mathematics (STEM) majors,” *Journal of College Science Teaching*, 2013. 42(5): p. 36-41.

16. P. M. Jansson, R.P. Ramachandran, J.L. Schmalzel, and S. Mandayam, “Creating an agile ECE learning environment through engineering clinics,” *IEEE Transactions on Education*, 2010. 53(3): p. 455-462.

17. R. F. DeMara, S. Salehi, N. Khoshavi, R. Hartshorne, and B. Chen, “Strengthening STEM Laboratory Assessment Using Student-Narrative Portfolios Interwoven with Online Evaluation,” in *Proceedings of American Association for Engineering Education Southeastern Conference & Exposition*, June 26 – 29, 2016, New Orleans, LA.

18. L. Katehi, G. Pearson, and M. Feder, Engineering in K-12 Education: Understanding the Status and Improving the Prospects. 2009: National Academies Press.

19. K. Squire and N. Patterson, Games and simulations in informal science settings, WCER: 2010.

20. X. Xiong and J.B. Beck, “Improving Long-Term Retention Level in an Environment of Personalized Expanding Intervals,” in *Proceedings of the 8th International Conference on Educational Data Mining*, June 26 – 29, 2015, Madrid, Spain, p. 582 – 583.

21. X. Xiong and J.E. Beck, “A Study of Exploring Different Schedules of Spacing and Retrieval Interval on Mathematics Skills in ITS Environment,” in *Proceedings of the International Conference on Intelligent Tutoring Systems*. June 5 – 9, 2014. Honolulu, HI, p. 504 – 509.

22. X. Xiong, Y. Wang, and J. Beck, “Improving students' long-term retention performance: a study on personalized retention schedules,” in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. Poughkeepsie, NY, March 16 - 20, 2015. p. 325-329.

23. N.T. Heffernan and C.L. Heffernan, “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*, 2014. 24(4): p. 470-497.

24. L. Razzaq and N.T. Heffernan, “Scaffolding vs. hints in the Assistment System,” in *Proceedings of International Conference on Intelligent Tutoring Systems*, June 26-30, 2006, Jhongli, Taiwan.

25. K. R. Koedinger and V. Aleven, “Exploring the assistance dilemma in experiments with cognitive tutors,” *Educational Psychology Review*, 2007. 19(3): p. 239-264.

26. I. I. Salame and S. A. Bacchus, “Assessment of the integration of online homework into general chemistry and its impact on student learning,” in *Abstracts of Papers of The American Chemical Society*. 2010. American Chemical Society, Washington, DC.

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

28. J.M. Webb, W.A. Stock, and M.T. McCarthy, “The effects of feedback timing on learning facts: The role of response confidence,” *Contemporary Educational Psychology*, 1994. 19(3): p. 251-265.
29. J. A. Kulik and C.-L.C. Kulik, “Timing of feedback and verbal learning,” *Review of educational research*, 1988. 58(1): p. 79-97.
30. H. L. Roediger and J. D. Karpicke, “Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention,” *Psychological Science*, 2006. 17(3): p. 249-255.
31. Framework for 21st century skills. *P21 - The Partnership for 21st Century Skills 2016*. Available from: http://www.p21.org/our-work/p21-framework.
32. A. S. Thomas, S. M. Bonner, H. T. Everson, and J. A. Somers, “Leveraging the power of peer-led learning: investigating effects on STEM performance in urban high schools,” *Educational Research and Evaluation*, 2015. 21(7-8): p. 537-557.
33. M. E. Muuro, R. O. Oboko, and P. W. Wagacha, “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*, 2016. 17(2).
34. D. Turgut, L. Massi, N. H. Bidoki, and S. S. Bacanli, “Multidisciplinary Undergraduate Research Experience in the Internet of Things: Student Outcomes, Faculty Perceptions, and Lessons Learned,” In *Proceedings of American Society for Engineering Education Annual Conference &* Exposition, June 25 – 28, 2017, Columbus, OH.