ASEE'S VIRTUAL CONFERENCE
At Home with Engineering Education
JUNE 22 - 26, 2020 #ASEEVC
Paper ID #28546

Does the use of cumulative and practice tests further improve a blended STEM classroom?

Prof. Autar Kaw, University of South Florida

Autar Kaw is a professor of Mechanical Engineering at the University of South Florida. He is a recipient of the 2012 U.S. Professor of the Year Award (doctoral and research universities) from the Council for Advancement and Support of Education and the Carnegie Foundation for Advancement of Teaching. Professor Kaw's main scholarly interests are in engineering education research, adaptive, blended and flipped learning, open courseware development, and the state and future of higher education. Funded by National Science Foundation, under Professor Kaw's leadership, he and his colleagues from around the nation have developed, implemented, refined, and assessed online resources for an open courseware in Numerical Methods. This courseware annually receives 1,000,000+ page views, 2,000,000+ views of the YouTube lectures, and 90,000+ visitors to the "numerical methods guy" blog. He has written more than 100 refereed technical papers and his opinion editorials have appeared in the Tampa Bay Times, Tampa Tribune and Chronicle Vitae. His work has been covered/cited/quoted in many media outlets including Chronicle of Higher Education, Inside Higher Education, U.S. Congressional Record, WUSF, Florida Senate Resolution, ASEE Prism, and Voice of America.

Dr. Renee M Clark, University of Pittsburgh

Renee Clark is Research Assistant Professor of Industrial Engineering and Director of Assessment for the Engineering Education Research Center (EERC) in the Swanson School of Engineering, University of Pittsburgh. She conducts research on education projects that focus on active learning and engineering professional development. Current research includes the propagation of active learning throughout the Swanson School and the use of systematic reflection and metacognitive activities within coursework. She received the Ph.D. in Industrial Engineering from the University of Pittsburgh and the MS in Mechanical Engineering from Case Western. She has over 25 years of experience as an engineer and analyst in industry and academia. She completed her post-doctoral studies in engineering education at the University of Pittsburgh.

Does the use of cumulative and practice tests further improve a blended STEM classroom?

Introduction

Since 1987, the first author has been teaching a junior-level Mechanical Engineering Numerical Methods course at the University of South Florida. Just by the nature of the content and by intention, the course has been continually transformed. For example, the programming languages used in the course chronologically were changed from FORTRAN to Visual Basic to Maple and currently MATLAB. And while having used mostly a talk-and-chalk mode in the classroom in the last century, the course has been taught formally in active-learning modes of blended and flipped learning since 2003.

The blended modality in the course itself evolved over the years, and now approximately one-third to one-half of the class time is spent on active learning activities such as think-pair-share [1], conceptual exercises via handouts or clickers, in-class procedural exercises, and outlining of programming projects and applied exercises. Many of these exercises are collected for a grade in the class. Some applied exercises, though, are taken home by students and graded after submission, as are the completed programming projects. Because the active learning displaces the coverage of some content in the class, students are assigned online digital audiovisual lectures and textbook readings for the rest of the content on a topic. To ensure students are getting practice on the topics covered in class and ones they do on their own, automatically graded online quizzes with no more than three questions per lesson are assigned (there are 8 topics that form 30 lessons for the course). Students also review and are assessed on pre-requisite online content before class time; however, students are not expected to learn or interact with the new content before coming to class.

The blended class improved cognitive learning and the classroom environment over the traditional classroom. The first evidence of this was from a prototype study [2] on two of the eight topics of the course, where the score on a final examination of a multiple-choice nature improved by 15% ($p<0.008$), while student satisfaction improved in the areas of classroom presentations, reading assignments, and problem sets. In another study [3] conducted over the summers of 2002-2006, where traditional, flipped, blended, and self-study modalities were compared for the same prototype topics, flipped and blended modalities were associated with better performance compared to the traditional and self-study modalities. Over the years, as the prototype study was extended to include all topics of the Numerical Methods course, we were unable to continue making such comparisons, as reviewers of the proposals and papers insisted on using free-response questions in the final examination for assessing higher-level thinking skills. Also, the survey instruments were updated, and new ones were adopted. However, as an unpublished result, the final examination score of 54% for the mostly traditional class improved to 62% for the blended class ($d=0.8$; $p<=0.008$) in the final implementation of the all-multiple-choice-question examination.

Having established that the blended class was associated with improved cognitive learning over the traditional class in the Numerical Methods course is congruent with the classic meta-analysis [4] on active learning by Freeman et al., where active learning was associated with an improved student performance by an average effect size of d=0.47 over traditional lecturing. A question naturally occurred over time to the instructor – Would a flipped classroom improve cognitive learning and classroom environment over a blended course? This question was asked through a three-institution study [5] led by the first author, where flipped learning was compared with blended learning. The results were mixed. On a final examination, for lower order-thinking skills problems, no practically-significant difference was found at two institutions for flipped instruction, while for another institution, a large practically-significant difference was found in favor of blended instruction. For higher-order thinking skills problems, small effect sizes were found at two institutions in favor of blended instruction, whereas at another, flipped instruction was associated with a small positive effect size [6]. Such mixed results also were noted in a 2019 meta-analysis by Sparkes [7], where flipped learning did not significantly outperform blended learning. We chose blended learning over traditional lecturing as the control group for our NSF study [5] because we had already shown that the blended class was better than the traditional class. Also, the classic meta-analysis [4] on active learning by Freeman et al. categorically points out that "the results [of active learning improving student performance by d=0.47 over traditional lecturing] raise questions about the continued use of traditional lecturing as a control in research studies, especially when it benefits underrepresented minorities (URMs) such as females and disadvantaged groups."

While the blended modality was associated with improved cognitive and classroom environment over the traditional classroom for numerical methods, the question that remains is "Can we improve the blended class even further by using other evidence-based learning strategies not used so far in the classroom, or would there be a limited effect of these interventions?" Would additional learning strategies, specifically,
1) cumulative tests,
2) practice tests, and
3) assigning but not grading regular in-class and homework assignments
further improve the blended classroom?

Cumulative final examinations are quite common in STEM higher education. Although such examinations are overwhelming to students, they are beneficial for long-term retention and hence provide better for future courses where the content of the course is a pre-requisite [8]. What if the mid-term tests were made cumulative as opposed to being unit tests?

Practice tests are known to improve student performance as they get to demonstrate their conceptual and procedural knowledge, become familiar with the format of the test, and get encouraged to discuss difficult topics. Based on a meta-analysis [9] of 272 studies, the average effect size was found to be 0.74 over other learning conditions. What if practice tests were offered to the students before each of the cumulative tests?

In an earlier study for the same course [10], we had found that there was no statistical and little practical significant difference in student final examination performance when homework was "assigned and graded" versus when homework was "assigned but not graded." While it does not affect student performance, this strategy lowers any student anxiety of being graded on formative assessment, no matter how low the stakes. It also frees the teaching assistant and the instructor from grading and use the time saved to increase one-on-one contact with students through office hours and in-class work. This was the last of the three simultaneous interventions that were applied in the improved blended class.

Research Questions

The following are hence the two research questions asked in this study:
RQ1: Are there performance differences on the final examination (cognitive learning) in a numerical methods course between the blended learning (control group) and modified blended learning (experimental group)?
RQ2: Are there differences in the perceptions of the classroom environment in a numerical methods course between blended learning (control group) and modified blended learning (experimental group)?

Methods

To find the effect of the interventions in the blended classroom, two groups were compared. We call them to be the modified blended learning (experimental) group, hereby called MBLG, and the blended learning (control) group, hereby called BLG. The participants of both groups were students in a Numerical Methods course taught at the junior level. The control group BLG had 198 participants (out of 283 registered students) from the semesters of spring 2014, spring 2015, and spring 2017. The experimental group MBLG had 50 participants (out of 66 registered students) from the semester fall 2019. The students who did not participate in the study were those who declined or neglected to give consent or withdrew from the course. No distinction was made in how participants or nonparticipants were taught. Rather, data from the nonparticipants were simply not used in the reporting of the study.

The control group BLG consists of conducting the course in a blended manner. Since there are many definitions of such modality in the literature, it becomes imperative to clarify our definition for the study. For BLG, students are not expected to come prepared to class with course content except for the pre-requisites. An example would be familiarity with the concept of the first derivative, slope, tangent to a curve, and equation of a tangent line from the pre-requisite Differential Calculus course before attending the class for the topic of Newton-Raphson method of solving nonlinear equations. The preparation is ensured through an automatically graded online quiz that is algorithmic and is due a few hours before the starting time of the class. The quizzes are conducted through a learning management system. In a typical class, about one-half to two-thirds of the time is used for lecturing while the rest of it is dedicated to active learning exercises such as peer-to-peer learning, solving procedural problems, outlining solutions to open-ended problems that are ill-defined, may need assumptions and additional data from reliable sources. Because of the displacement of class time due to active learning, some content on a topic is pushed to out-of-class time to foster self-efficacy and life-long learning skills. The graded assessment includes weekly automatically graded online quizzes, two main projects, special assignments such as open-ended problems, four tests, and a final examination. Non-graded assignments include multiple-choice questions and selected problems from the textbook.

The experimental group MBLG is a modified version of the BLG. The MBLG is different only in the following three ways – 1) each of the tests is cumulative rather than a unit test, 2) practice tests are offered to the students, 3) none of the assigned homework was graded. Projects continued to be part of the graded assessment as they are critical in a course like Numerical Methods.

The first author gives 3- to 4-unit tests during the course. The cumulative examinations consist only of a concept inventory test and a final examination, and those are given at the end of the semester. The unit tests were, therefore, replaced by cumulative tests for the experimental condition (MLBG). Although the content emphasis on the cumulative test was on recently covered topics since the earlier test, the previous content was nonetheless a substantial part of the test, varying from 33% to 50% of the test by score.

Practice tests were given to students before each midterm test in the MLBG. To avoid any bias, the practice tests were made through a computer program written by the first author that selects problems randomly from a database consisting of 167 questions. In the computer program, one can input the percentage of questions sought from new versus previous content, and the percentage sought from short questions (fill-in-the-blank and multiple-choice) versus free-response questions. The questions in the database were the same that were used in the LBG and MLBG for in-class work and homework. For the first test, one practice test was posted, while two practice tests were posted for each of the next three tests. The practice tests were posted two days before the test to avoid 1) students preparing using only the practice tests, and 2) developing a false sense of preparation. All students were recommended to take the practice tests at home in conditions similar to the tests (i.e., use 75 minutes of time, a TI30Xa calculator, and an instructor-made formula sheet).

In the control group LBG, students were assigned both graded and non-graded homework. For example, they completed an automatically graded online quiz on every sub-topic. This quiz was assigned on Thursday after class and was due on Tuesday before class. A problem set from the textbook was also assigned every week but was not graded. In-class active learning exercises were also assigned, and only some were graded. In the experimental group MLBG, the same in-class exercises and problem sets were assigned but were not graded. Also, automatically-graded online quizzes were not assigned.

For both the groups, a final examination was used to measure student performance, while a survey was conducted to measure the classroom environment.

Results

Cognitive learning and classroom environments were measured to compare the modified blended learning (experimental) group (MBLG) and the blended learning (control) group (BLG).

Cognitive Learning

To compare cognitive learning, we used a 2-hour final examination that is given during the last week of the semester. The final examination was identical for both groups. The exam has 14 multiple-choice (lower-level thinking) and 4 free-response (higher-level thinking) questions.

Each of these two parts of the final exam is equally weighted. The free-response questions are graded on a scale of 1-4 using a holistic rubric, as given below.

4 – The student demonstrates a complete understanding of the problem. All requirements of the task are included in the response.
3 – The student demonstrates a considerable understanding of the problem. All requirements of the task are included.
2 – The student demonstrates a partial understanding of the problem. Most requirements of the task are included.
1 – The student demonstrates little understanding of the problem. Many requirements of the task are missing (Is there anything salvageable in the given solution?)
0 – The student leaves the solution blank or shows no demonstration of understanding of the problem – simply repeats the data or copies formulas, or the student shows or uses formulas that are not even relevant.

The difference in the average student score on the final examination in the two treatments was examined by using an analysis of covariance (ANCOVA), with GPA from the pre-requisite coursework serving as the control variable. The difference in the average score was not statistically significant (p=0.165), and the effect size was Cohen's d=0.27 in favor of the modified blended classroom (see Table 1). Effect size [11] quantifies the difference between an experimental and a control group and is approximately defined as (Mean of the experimental group–Mean of the control group)/(standard deviation). To appreciate the effect size of d=0.27, values of effect sizes obtained should be compared to comparable studies [12] in the educational interventions field where the average effect size is d=0.38 for published research [13] and d=0.18 for unpublished research [14] are reported.

Table 1. Final Examination Performance Comparison

	Blended Learning (Control)	**Modified Blended Learning (Experimental)**		
	Average Percentage (Standard Deviation)		**p-value (two-tailed)**	**Effect size**
Final Examination Score	49.73 (15.96)	54.04 (16.01)	0.165	0.27
Sample Size, n	188	50		

<u>Classroom Environments</u>

Several studies show a correlation between learning environments and student academic achievement [15-17]. The discernment of the students learning environment is mostly measured by surveys, and one such survey is the well-established College and University Classroom Environment Inventory (CUCEI) [15]. The inventory was first developed in 1987 and is sufficiently validated and reliable – scale alpha reliabilities range from 0.70 to 0.90 [16], while overall Cronbach reliabilities as high as 0.91 have been reported [18]. Fraser, et al [17] found that satisfaction in higher education classes is associated with all the seven psychosocial dimensions (Table 2) that the CUCEI measures.

Table 2. Seven psychosocial dimensions of CUCEI inventory [15]

Psychosocial Dimension	**Extent to Which**
Cohesiveness	Students know and help one another
Individualization	Students are treated individually and differentially
Innovation	New class activities or teaching techniques are used
Task orientation	Class activities are well-organized
Involvement	Students participate in class activities
Personalization	Interaction takes place with instructor and there is concern for students
Satisfaction	Classes are enjoyed by students

The inventory consists of 49 questions on the above seven dimensions, and students rate the questions on a scale of 1 (strongly disagree) to 5 (strongly agree).

For our study, the response rate was 70% for students enrolled in the BLG and 76% for students enrolled in the MLBG. The two instructional methods were compared using *t*-tests, assuming unequal variances over the seven dimensions that make up the CUCEI classroom environment inventory. The effect sizes for all seven dimensions were calculated by using a multivariate analysis of covariance (MANCOVA), and they ranged from -0.19 to -0.98, all in favor of the blended classroom (LBG). The differences were generally statistically significant, as shown in Table 3. The authors do not have a categorical explanation, but it may be because cumulative tests are not looked upon favorably by students [8]. Also, since homework assignments were not graded, many students may not have been as prepared for the in-class active learning or the pre-requisite knowledge needed for the new content covered in the lectures. In the spring 2020 semester, we are hence reverting to assigning graded online quizzes to improve pre-class preparation and possibly the classroom environment likely.

Discussion, Conclusions, and Limitations

Active learning in blended classrooms has already been associated with improved student performance. To further improve the cognitive learning outcomes and the class environment in a blended classroom, unit tests were replaced by cumulative tests, practice tests were offered, and homework assignments were assigned but not graded.

The modified blended classroom showed a small improvement in cognitive performance as measured through a final examination. An effect size of d=0.27 was found, which is comparable to other educational interventions. The classroom environment, however, was reported to be less favorable in the modified blended classroom. This may be due to how cumulative tests are viewed by students. The instructor will continue to give cumulative tests and assign weekly online homework. The latter, although of a low-stakes nature, would likely compel students to be prepared for the in-class assignments as well as with the pre-requisite knowledge needed for better comprehension of new content during the classroom lecture.

Table 3. Classroom Environment Comparison

	Blended (Control)	Modified Blended (Experimental)		
Dimension	**Average on Scale of 1-5 (Standard Deviation)**		**p-value (two-tailed)**	**Effect size**
Cohesiveness	2.97 (0.80)	2.81 (0.79)	0.228	– 0.19
Individualization	2.49 (0.64)	2.30 (0.61)	0.073	– 0.29
Innovation	3.19 (0.64)	2.79 (0.63)	<0.001	– 0.62
Involvement	3.31 (0.60)	3.01 (0.69)	0.009	– 0.48
Personalization	4.08 (0.62)	3.65 (0.75)	0.001	– 0.66
Satisfaction	3.53 (0.92)	3.01 (1.12)	0.005	– 0.53
Task orientation	4.17 (0.54)	3.59 (0.74)	<0.001	– 0.98
Sample Size, n	198	47		

Relative to study limitations, although the sample size in the experimental group was smaller than in the control group, it was nonetheless sufficient for the statistical analysis performed. As the experimental approach is continued in future semesters of the course (including spring 2020), additional data will be collected and analyzed. To account for differences in student academic performance history between the experimental and control groups, an analysis of covariance (ANCOVA) was used to compare the exam results, with GPA from the pre-requisite coursework serving as the control variable.

This article adds to the authors' notable publications on the blended instructional approach in a STEM classroom, in particular, by investigating various evidence-based learning strategies within this type of classroom format.

References

[1] F. Lyman, "Think-Pair-Share: An Expanding Teaching Technique," *Maa-Cie Cooperative News,* vol. 1, no. 1, pp. 1-2, 1987.

[2] A. Kaw, G. H. Besterfield, and J. Eison, "Assessment of a Web-Enhanced Course in Numerical Methods," *International Journal of Engineering Education,* vol. 21, no. 4, pp. 712-722, 2005.

[3] A. Kaw and M. Hess, "Comparing Effectiveness of Instructional Delivery Modalities in an Engineering Course," *International Journal of Engineering Education,* vol. 23, no. 3, pp. 508-516, 2007.

[4] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, "Active Learning Increases Student Performance in Science, Engineering, and Mathematics," *Proceedings of the National Academy of Sciences,* vol. 111, no. 23, pp. 8410-8415, 2014.

[5] A. Kaw, M. Besterfield-Sacre, Y. Lou, and A. Scott, "Improving and Assessing Student Learning in an Inverted Stem Classroom Setting," ed: http://nsf.gov/awardsearch/showAward?AWD_ID=1322586 (accessed on January 24, 2020).

[6] R. M. Clark, A. Kaw, Y. Lou, A. Scott, and M. Besterfield-Sacre, "Evaluating Blended and Flipped Instruction in Numerical Methods at Multiple Engineering Schools," *International Journal for the Scholarship of Teaching & Learning,* vol. 12, no. 1, 2018.

[7] C. Sparkes, "Flipped Classrooms Versus Traditional Classrooms: A Systematic Review and Meta-Analysis of Student Achievement in Higher Education," Ph.D. Dissertation, Educational Technology, Concordia University, 2019. [Online]. Available: https://spectrum.library.concordia.ca/985276/

[8] S. S. Sansgiry, A. Nadkarni, and T. Lemke, "Test Anxiety with Respect to a Comprehensive Cumulative Assessment," *Journal of Pharmacy Teaching,* vol. 12, no. 1, pp. 41-59, 2005.

[9] O. O. Adesope, D. A. Trevisan, and N. Sundararajan, "Rethinking the Use of Tests: A Meta-Analysis of Practice Testing," *Review of Educational Research,* vol. 87, no. 3, pp. 659-701, 2017.

[10] A. Yalcin and A. Kaw, "Does Grading Homework Improve Student Examination Performance?," *International Journal of Engineering Education,* vol. 27, no. 6, pp. 1333-1342, 2011.

[11] P. D. Ellis, *The Essential Guide to Effect Sizes: An Introduction to Statistical Power.* Cambridge University Press, 2010.

[12] J. Cohen, *Statistical Power Analysis for the Social Sciences*. Lawrence Earlbaum Associates, 1988.

[13] J. Hattie, *Visible Learning*. Oxfordshire, Routledge, 2008.

[14] J. Polanin, E. Tanner-Smith, and E. Hennessy, "Estimating the Difference between Published and Unpublished Effect Sizes a Meta-Review," *Review of Educational Research,* vol. 86, no. 1, pp. 207-236, 2016.

[15] "College and University Classroom Environment Inventory." https://case.edu/ucite/sites/case.edu.ucite/files/2018-02/College-and-University-Classroom-Environment-Inventory.pdf (accessed November 25, 2019).

[16] T. Byers, M. Mahat, K. Liu, A. Knock, W. Imms. *Systematic Review of the Effects of Learning Environments on Student Learning Outcomes*, University of Melbourne Technical Report, 2018, http://www.iletc.com.au/wp-content/uploads/2018/07/TR4_Web.pdf (accessed March 4, 2020).
[17] B. Fraser, D. Treagust. "Validity and use of an instrument for assessing classroom psychosocial environment in higher education." *Higher Education*, vol. 15, pp. 37–57, 1986.
[18] M.Chiu, H. Tuan, H. Wu, J. Lin, C. Chou, *Chemistry Education and Sustainability*, Springer, NY, 2013.