International Journal of Engineering Education Vol. 37, No. 3, pp. 585–593, 2021
Printed in Great Britain
0949-149X/91 $3.00+0.00

Enhancing Student Outcomes in a Blended Numerical Methods Course for Engineers: The Case for Practice and Cumulative Tests*

RENEE M. CLARK[1] and AUTAR KAW[2]
[1] Industrial Engineering, University of Pittsburgh, 1025 Benedum Hall, Pittsburgh, PA 15261, USA. E-mail: rmclark@pitt.edu
[2] Department of Mechanical Engineering, University of South Florida, 4202 East Fowler Avenue ENB118, Tampa, FL 33620-5350, USA. E-mail: kaw@usf.edu

Evidence-based testing strategies in the form of multiple cumulative midterm tests preceded by practice tests were recently introduced to a numerical methods course for engineers after the course had been taught for many years in a blended fashion. The instructor introduced these practices in fall 2019, thereby creating his so-called modified blended approach, with the objective of enhancing direct and affective assessment results in his blended classroom implementation. A statistical comparison of results from this modified approach with results from a prior semester of blended instruction was made using final exam and concept inventory scores as well as classroom environment scores based on the CUCEI. This comparison was made for students collectively and for several demographic segments of interest. Based on triangulated results from direct assessments of conceptual understanding and Bloom's taxonomy (lower levels), the modified blended approach with the testing strategies may be the preferred method for this blended classroom for students collectively as well as potentially for Pell grant recipients as a group. The classroom environment and direct assessment results from the higher levels of Bloom's taxonomy did not suggest a preferred instructional method. Support for blended instruction and practice and cumulative testing from the literature is also presented.

Keywords: blended classroom; numerical methods; cumulative test; practice test

1. Introduction and Literature Review

For over 30 years, the second author has continually transformed his undergraduate numerical methods course for engineers at a large university in the southern U.S. This course encompasses numerical methods for solving nonlinear equations, simultaneous linear equations, regression, integration, scientific computing, differentiation, interpolation, and ordinary differential equations. While having used a talk-and-chalk style before 2000, the instructor began formally using and studying active-learning approaches in his classroom in 2003, including both blended and flipped methodologies and combinations of them. His approaches to blended and flipped instruction within this required mechanical engineering course naturally evolved over time through various funded research, including the integration of adaptive software for pre-class preparation in the flipped classroom [1–4]. Several publications describe outcomes since 2003 from the instructor's use of blended instruction, including the notable outcome that blended (as well as flipped) instruction was associated with better student performance than traditional (i.e., lecture) and self-study approaches [5, 6]. Given very strong interest in continual course improvement and engineering education scholarship, many ideas and research questions naturally occurred to the instructor over time for further enhancement of his blended classroom, including the use of evidence-based testing strategies. As an avid practitioner of technology-enhanced instruction, active learning, and continual course transformation, he developed an interest in researching the effect of these testing practices in his blended STEM classroom. He specifically began to ask, "*Can the blended classroom be improved relative to student outcomes by using additional evidence-based testing strategies*"*?* Specifically, the instructor developed an interest in introducing cumulative midterm tests preceded by practice tests, in line with positive findings from the educational psychology literature, as will be detailed in an upcoming section. He, therefore, introduced a new testing approach in fall 2019 for a blended-classroom offering of his numerical methods course, in essence, creating a modified blended classroom. One aspect of this approach was outside-the-classroom practice tests posted before the in-class cumulative midterm tests. Subsequent comparison of direct and affective student outcomes from this modified blended offering to those from a prior blended semester was made. Based on direct assessment measures, there is evidence that the modified blended classroom may be the favoured approach. The next section begins with a description of and support for blended instruction based on the literature, fol-

* Accepted 13 November 2020.

lowed by a review of and support for practice and cumulative testing strategies from the psychology literature.

1.1 Literature Review: Blended Instruction

Blended instruction replaces a portion of face-to-face classroom instruction with online or technology-driven resources while typically still maintaining the traditional classroom format [7–9]. Blended instruction can be conceived as a combination, or blending, of two or more learning methods or media that can create a learning environment and tools that are responsive to individual student differences and enable autonomous learning [10]. Online or technology-driven educational resources may include simulations, labs, video tutorials, and assessments [9]. Blended learning entails using the web and class time according to what each does best [11]. Blended practices were the subject of an instructional redesign program funded by the Pew Charitable Trusts, in which institutions of higher education redesigned their instruction using technology, including computer-based assessments, online discussion groups/learning communities, and online tutorials [12, 13]. Blended instruction in other mechanical and electrical engineering courses has entailed online experiments, labs, and simulations [14–20]. In comparisons of blended versus traditional instruction, blended implementations have shown success. For example, in round one of the Pew Charitable Trusts program discussed above, five projects (out of ten) reported improved outcomes, and four reported equivalent outcomes [12]. Further, in a multi-semester comparison of face-to-face, fully online, and blended instruction, the last had the highest percentage of students who earned a "C" or greater [21].

1.2 Literature Review: Testing Strategies – Practice and Cumulative Testing

Tests are likely unwelcome by many students [22]. However, learning is promoted when students take a test, and this finding is often overlooked in education [22, 23]. There are multiple specific reasons from a learning perspective as to why increased testing or retrieval practices should be done in education, including retention of information, formative identification of knowledge gaps, metacognitive self-monitoring, encouragement to study, and more-transferrable learning [24, 25]. Compared to rereading material, correctly retrieving material from memory can have a direct impact on long-term memory [26]. The indirect, self-regulatory impacts on learning, mentioned above can result from a student better deciding what he/she must re-study or practice more, or from a student being able to determine what he/she knows or does not know [26]. Over 100 years of study has shown that practice tests boost retention and learning compared to simply rereading material, where a practice test is defined as a minimal-stakes activity completed outside the classroom [22, 26].

Thus, this so-called *testing effect* is associated with greater learning and retention of material given the retrieval processes involved, as discussed by additional articles [23, 25, 27]. Students who have taken practice tests on the material before taking a final exam on the same material often performed better than students who used other methods for studying, such as re-studying or different non-test approaches [28]. Specifically, a recent meta-analysis of 118 articles showed practice tests to be associated with an average medium effect size of 0.51 compared to re-studying the material [28]. It was found that for delayed assessment (i.e., 2 to 7 days after content delivery), having taken a practice test was associated with greater content retention than having re-studied the material [27].

Cumulative examinations, particularly high-stakes summative exams, can certainly be undesirable to and anxiety-producing for students [22, 26, 29]. However, in their article "*Expectation of a final cumulative test enhanced long-term retention*," Szpunar and colleagues found that not only did taking an initial test improve performance on the final test (relative to not taking an initial test), but establishing the expectation of a final cumulative exam enhanced achievement on this exam, versus if that expectation had not been set [30]. The authors surmised that if students are expecting a cumulative exam, they may integrate information across units and, in essence, study more effectively [30]. Thus, it is plausible that because these students had to remember information for a later exam, they were encouraged to process the material continually and positively impact the availability of this material during later recall [30]. This is in contrast to those students who do not work under this expectation, as they may disregard any prior information and focus just on the new information they will be tested on [30]. Alternatively, knowledge of a cumulative exam may encourage students to study or re-study the course material in a distributed fashion over time (versus cramming the night before the exam), with distributed practice and study resulting in longer-term retention of knowledge and skills [26]. In summary, the evidence for cumulative midterm and final exams with regards to deep learning and persistent retention is exceptionally strong, keeping in mind that these types of assessments require continued and ongoing interaction by students with the course content and thereby increased chances for understanding and retention [31]. For long-term retention of material and end-of-course performance, cumulative exams given throughout the course (i.e., cumulative midterms) are the frontrunner testing approach based on numerous research studies [31].

In an introductory psychology course, students who took multiple cumulative unit tests had better long-term retention on the final exam versus those who took a series of non-cumulative unit exams, as evidenced by a statistically and practically significant final exam score [32]. In a second study with introductory psychology coursework, a significantly-higher score was found on a departmental post-course assessment for the cohort who had taken a cumulative (versus non-cumulative) final exam [33]. In conclusion, given this background and support from the literature related to practice and cumulative testing approaches, the research questions evolved to the following:

RQ1: What is the extent of the performance difference on lower-order-cognition final-exam questions when practice and cumulative tests are used in the blended classroom? What is the extent of this difference for particular demographic groups, including underrepresented minorities and Pell Grant recipients?

RQ2: What is the extent of the difference in conceptual understanding of course material when practice and cumulative tests are used in the blended classroom? What is the extent of this difference for particular demographic groups, including underrepresented minorities and Pell Grant recipients?

RQ3: What is the extent of the performance difference on higher-order-cognition final-exam questions when practice and cumulative tests are used in the blended classroom? What is the extent of this difference for particular demographic groups, including underrepresented minorities and Pell Grant recipients?

RQ4: Is there a difference in the perception of the classroom environment when practice and cumulative tests are used in the blended classroom?

2. Methods

2.1 Instructional Methods

The instructor has been teaching the course for over 30 years using various approaches, including flipped, flipped with adaptive learning, and blended instruction. With the blended approach, although students had access to videos, they were not required nor expected to watch them before class, except in the case of the pre-requisite content, for which they were assessed online. Class time was spent on the presentation of new material, clicker questions (non-graded) with peer-to-peer and instructor-led discussions, and solution planning for applied problems. Specifically, approximately 33–50% of class time was spent on active learning activities, including think-pair-share, conceptual exercises via worksheets or clickers, procedural exercises, outlining of programming projects and applied exercises, and discussion. Many of the in-class exercises were collected for grading both at the end of class and after completion at home. Since the active learning replaced some content covered during class, video lectures and textbook readings were assigned for some content. To drive accountability and practice, automatically-graded online quizzes were assigned. During the spring 2017 semester with a blended classroom, a newly-completed Numerical Methods concept inventory was used in the class, along with the instructor's traditional cumulative final exam. For this reason, the spring 2017 semester served as the comparison (i.e., control) semester. Each served as a direct assessment of student learning of numerical methods course content.

After several semesters during fall 2019, the course was again taught in a blended format, but with the addition of practice tests and four cumulative midterm exams involving multiple units of material to determine if these strategies were associated with enhanced direct and affective outcomes in this modified blended classroom. Course content is often taught in the form of units, each followed by a unit exam [30]. For each cumulative midterm exam, the emphasis was on recently-covered content, but previous material still comprised a significant portion of the exam (i.e., 33% to 50% of the exam score).

One or two non-graded practice tests were posted before each midterm exam and were developed using a computer program that selected problems randomly from a database of over 150 questions. Random generation of problems was done to avoid bias, and the instructor was able to input the percentage of questions desired from new content versus prior content. The practice tests were posted to the learning management system two days beforehand to avoid student preparation using just the practice tests. Students took the practice tests outside the classroom and could submit their responses to the instructor for feedback. Students were informed via the syllabus and on the first day of class that there would be a series of cumulative midterm tests and a cumulative final at the end of the semester. The high-level research question became the following: "*Can blended-classroom student outcomes be enhanced by employing the evidence-based strategies of cumulative and practice tests, or is there a limited effect of these strategies?*"

2.2 Assessment Methods: Comparison of Final Exam, Concept Inventory, and Classroom Environment Results

In designing the assessment plan for this study, the approach used by the authors in previous NSF-funded studies on blended versus flipped instruction as well as adaptive preparation for the flipped classroom was adopted [1–3]. Therefore, a triangulated-data approach consisting of comparisons involving a final exam, concept inventory, and classroom environment inventory was used. Specifically, the College and University Classroom Environment Inventory (CUCEI) was used to assess student perceptions on seven psychosocial dimensions of the classroom [34]. Examples of psychosocial dimensions from the CUCEI include Individualization (i.e., individual/ differential treatment of students and student decision-making power), Innovation (i.e., novel and unusual class activities or teaching techniques), and Involvement (i.e., active student participation in class activities). There are seven questions per dimension, each on a 1 to 5 scale, with 5 being most desirable for the environment of the classroom.

A demographics survey enabled the collection of data for investigating student groups of interest, in particular underrepresented minority students (URM) and Pell grant recipients. Although data on the student's gender was collected, small sample sizes precluded statistical analysis by gender. The final exam and concept inventory scores were compared in a stratified fashion to investigate potential differences for the URM and Pell grant students between the two instructional approaches. The demographics survey was also used to collect data on grades achieved by the students in the pre-requisite coursework, such as calculus 1–3 and ordinary differential equations. The pre-requisite GPA served as the control variable for the analysis of covariance (ANCOVA) comparisons of final exam and concept inventory scores between the blended and modified blended approaches. The use of a control variable such as GPA as part of an ANCOVA analysis accounts for any differences in academic performance history between two or more student cohorts [35].

The final exam and concept inventory remained exactly the same for the two instructional approaches and were similar in that they each contained multiple-choice questions requiring little to no calculations. The multiple-choice questions on the final exam were mainly based on the lower levels of Bloom's taxonomy (i.e., remember, understand, apply) and were evenly distributed across all eight topics of the course. The free-response problems on the final exam required more open-ended analysis and targeted the higher levels of Bloom's taxonomy, such as analysis or evaluation. A holistic rubric with a 0–4 rating scale was used by the instructor for consistently grading the free-response questions. The scale was defined as follows: (0): no understanding; (1): little understanding, with many requirements missing; (2): partial understanding, with most requirements completed; (3) considerable understanding, with all requirements completed; and (4) complete understanding, with all requirements completed.

The concept inventory measured the student's conceptual understanding, focused on critical thinking and logic, and highlighted any misconceptions [36]. The inventory encompassed six key concepts that were associated with the greatest number of misconceptions, as agreed upon by a development team through a Delphi process [36].

Given the smaller sample sizes associated with the URM and Pell grant demographic groups and the accompanying uncertainty of normally-distributed scores, the non-parametric version of analysis of covariance, known as Quade's Test, was reported for this study [37–38]. However, the non-parametric results coincided with those from the parametric version of ANCOVA, meaning that the p-values from the parametric and non-parametric tests were in agreement [35].

Cohen's d or Hedge's g effect sizes were also calculated to measure the magnitude and practical significance of the difference between each pair of means, with Hedge's g used in the case of smaller sample sizes [39–41]. When calculating the effect sizes, the blended method was considered the reference (i.e., control) group relative to the modified blended approach.

3. Results

3.1 Direct Assessment of Learning: Final Exam and Concept Inventory Comparisons

To directly assess and compare student performance in the blended versus modified blended classrooms (i.e., the first and second research questions), final exam and concept inventory results were statistically analysed. For the final exam, students' multiple-choice and free-response performance were examined separately. The statistical analysis was done for participating students as a whole as well as the URM and Pell-grant-recipient groups, as shown in Table 1. The students taking the course were junior-level mechanical engineering students. For the direct-assessment statistical analysis, there was demographic and corresponding exam/concept inventory data for 57% of the enrolled students during the spring 2017 blended version of the course, and for the fall 2019 modified blended version, this percentage was 78%. These percentages are influenced by student interest in and subsequent participation in the research study, including actual completion of the various key instruments, such as the demographics survey. The corresponding sample sizes available for statistical analysis were as follows: 62 students during the spring 2017 semester and 49 students during fall 2019. Although the voluntary participation rate was relatively lower for the spring 2017 blended classroom, it still constituted a majority of the students enrolled.

3.1.1 Research Question 1: Lower-Order Cognition – Final Exam Scores

For students as a whole, the modified blended classroom was associated with a significantly higher average adjusted multiple-choice final exam score (65.2% vs. 57.9%; $p = 0.009$) after applying the Bonferroni adjustment for multiple comparisons [42]. Since three tests were run for the three demographic groups in Table 1, each p-value from the ANCOVA test was multiplied by three, as given in the *p adj* column of Table 1. Since the sample sizes were small for the demographic segments, Hedge's g, which is very similar to Cohen's d, was reported as the effect size within Table 1. The effect size associated with all students was medium at Hedge's $g = 0.51$. For the URM demographic group, the difference in the two instructional methods was not significantly different from zero, although for the Pell grant recipients, it was after applying the Bonferroni adjustment ($p = 0.036$). For the Pell Grant recipients, the effect size was approximately large at $g = 0.78$. Thus, for students as a whole and for the Pell Grant recipients, the modified blended classroom was the favoured instructional approach based upon the multiple-choice outcomes.

3.1.2 Research Question 2: Conceptual Understanding – Concept Inventory Scores

With the concept inventory, the outcomes were similar (to those obtained with the lower-order final-exam questions) for students as a whole. That is, the modified blended approach was associated with a significantly higher average adjusted score (60.6% vs. 51.6%; $p = 0.009$) after applying the Bonferroni adjustment for multiple comparisons, as given in Table 2. The effect size was also medium at $d = 0.59$. However, for the URM and Pell grant groups, the two instructional methods were not associated with significantly different concept inventory scores, although the effect sizes were each approximately medium. Thus, there is evidence that the modified blended classroom may have been the favoured instructional approach based upon the concept inventory outcomes as

Table 1. Final Exam Multiple Choice Comparison: Blended vs. Modified Blended

Dem Group	Adjusted Mean Percentage % (*s*) n		Quade's Test *p unadj*	Quade's Test *p adj*	Effect Size *g*
	Blend	**Mod Blend**			
All	**57.9** *(14.3)* 62	**65.2** *(14.3)* 50	0.003	0.009	0.51
URM	**54.9** *(14.7)* 17	**60.6** *(14.8)* 13	0.290	0.870	0.38
Pell	**56.7** *(14.5)* 21	**68.2** *(14.6)* 17	0.012	0.036	0.78

The blended method is the reference (i.e., control) when comparing it to the modified blended.

Table 2. Concept Inventory Comparison: Blended vs. Modified Blended

Dem Group	Adjusted Mean Percentage % (*s*) n		Quade's Test *p unadj*	Quade's Test *p adj*	Effect Size *g*
	Blend	**Mod Blend**			
All	**51.6** *(15.2)* 62	**60.6** *(15.3)* 49	0.003	0.009	0.59
URM	**47.4** *(16.2)* 17	**55.2** *(16.3)* 13	0.119	0.357	0.48
Pell	**54.7** *(15.1)* 21	**62.2** *(15.1)* 17	0.119	0.357	0.49

The blended method is the reference (i.e., control) when comparing it to the modified blended.

Table 3. Final Exam Free Response Comparison: Blended vs. Modified Blended

Dem Group	Adjusted Mean Percentage % (*s*) n		Quade's Test *p* *unadj*	Quade's Test *p* *adj*	Effect Size *g*
	Blend	**Mod Blend**			
All	**41.6** *(19.4)* 62	**39.9** *(19.5)* 50	0.634	1.000	–0.09
URM	**37.0** *(20.9)* 17	**41.0** *(20.9)* 13	0.599	1.000	0.18
Pell	**45.7** *(22.2)* 21	**41.7** *(22.3)* 17	0.700	1.000	–0.18

The blended method is the reference (i.e., control) when comparing it to the modified blended.

well. As discussed in the Methods section, the concept inventory consisted of multiple-choice questions that required minimal calculations. Therefore, it was similar to the multiple-choice section of the final exam, serving to triangulate the favourable results for the modified blended classroom to a certain degree.

3.1.3 Research Question 3: Higher-Order Cognition – Final Exam Scores

Interestingly, such patterns were not found with the free-response question outcomes on the final exam (i.e., the third research question). The adjusted average percentages were not significantly different between the blended and modified blended classrooms for any of the demographic groups, and the effect sizes were small, as shown in Table 3. Thus, there was not a preferred instructional approach in this study with respect to the more open-ended problems students were asked to solve.

3.2 Affective Assessment: Classroom Environment – Research Question 4

To investigate the final research question about comparing the perceived classroom environment among the two instructional methods, the CUCEI, or College and University Classroom Environment Inventory, was employed. For statistical analysis of the affective data on the classroom environment, there were responses from 58% of enrolled students in the spring 2017 semester (i.e., blended classroom) and 75% of enrolled students during the fall 2019 semester (i.e., modified blended classroom). The corresponding sample sizes available for this statistical analysis were 63 students during spring 2017 and 47 students during fall 2019.

The instructional methods were compared using a MANOVA, or multivariate analysis of variance, since there were seven outcome variables (i.e., dependent variables) corresponding to the seven dimensions of the CUCEI classroom environment inventory, as shown in Table 4 [35, 43]. The mean of each dimension (i.e., 1 to 5 scale) and standard deviation are shown in Table 4. Given the larger sample sizes (compared to the demographic-based analyses), Cohen's *d* effect sizes were calculated, also shown in Table 4. The omnibus test indicated a significant effect of the instructional method on the classroom environment ($p < 0.014$). The higher CUCEI scores were associated with the blended classroom, with six of the seven dimensions having a higher average value in the blended classroom versus the modified blended classroom. However, the univariate ANOVA tests indicated a significant difference only for one CUCEI dimension (i.e., Task Orientation; $p = 0.009$), and this difference did not remain significant after applying the Bonferroni correction for multiple comparisons by multiplying each univariate *p*-value by 7. For this dimension, the effect size was medium at $d = –0.51$, with the blended classroom being the reference category. All other effect sizes were of small magnitude.

From a summary perspective, although the blended classroom received the higher classroom environment scores, the differences with the modified blended classroom were not statistically significant, and only one effect size was medium, with the others being small. The lack of a definitive preference for our modified blended classroom, in particular, relative to the Innovation, Personalization, and Satisfaction dimensions, may be related to students' negative view of cumulative exams and tests in general and the fact there were several of them, despite this approach being associated with enhanced academic performance. Therefore, a definitively-favoured classroom environment did not emerge when comparing the blended and modified blended classrooms.

4. Discussion

This article investigated evidence-based testing strategies in the form of multiple cumulative mid-

Table 4. Classroom Environment Comparison: Blended vs. Modified Blended

Dim	Mean (s)		Univar p *unadj*	Univar p *adj*	Effect Size d
	Blend	Mod Blend			
Coh	2.62 *(0.69)*	2.82 *(0.79)*	0.159	1.000	0.27
Indiv	2.48 *(0.62)*	2.31 *(0.62)*	0.160	1.000	–0.27
Inn	2.97 *(0.69)*	2.80 *(0.63)*	0.179	1.000	–0.26
Invol	3.02 *(0.58)*	3.02 *(0.70)*	0.966	1.000	–0.01
Pers	3.88 *(0.70)*	3.65 *(0.75)*	0.109	0.763	–0.31
Satis	3.08 *(0.99)*	3.02 *(1.12)*	0.754	1.000	–0.06
Task Or	3.94 *(0.63)*	3.60 *(0.74)*	0.009	0.063	–0.51
n	63	47			

Coh = Cohesiveness (Students know & help one another).
Indiv = Individualization (Students treated individually/differentially & can make decisions).
Inn = Innovation (Novel class activities or teaching techniques).
Invol = Involvement (Active student participation in class activities).
Pers = Personalization (Interaction w/ instructor & concern for student welfare).
Satis = Satisfaction (Enjoyment of classes).
Task Or = Task orientation (Organization and clarity of class activities).

term tests preceded by practice tests within a blended STEM classroom. This work makes an addition to the authors' research on blended and other enhanced instructional approaches within STEM education. An assessment plan to directly measure student learning as well as student perspectives and feelings was utilized. The plan consisted of two established instruments for direct assessment of learning as well as the College and University Classroom Environment Inventory for affective assessment.

Based on the instructor's assessment of the performance outcomes in his course as well as support from the psychology literature in general, he plans to continue to employ cumulative midterm exams preceded by practice exams. The instructor believes it's also a good approach in online or remote courses, especially in the age of COVID-19, as it decreases the ability of students to merely sift through their notes to find the correct answer or approach to a similar problem within a limited time span, given the larger volume of material they are expected to be familiar with.

Given that engineering students from just one public U.S. university were studied, the generalizability of our results may be somewhat limited. Students were not randomly assigned to the two treatment groups, and so this study was quasi-experimental, as are many educational studies. To account for a student's academic history as a possible confounding factor, the pre-requisite GPA was used as a control variable within the analysis of covariance testing of the two treatments. The sample sizes for the specific demographic segments were small, reducing power to detect differences. However, conservative statistical procedures (i.e., non-parametric testing and effect sizes) were used. A further limitation of this research is related to the use of exams as outcomes, as they are relatively short-term performance measures compared to outcomes that may be apparent several years into the students' careers.

5. Conclusions

Based on multiple-choice outcomes from both the cumulative final exam and concept inventory (CI), the modified blended classroom was the favoured approach, having significantly higher scores (p = 0.009) and medium effect sizes of g = 0.51 and g = 0.59, respectively. In addition, there was some evidence that the modified blended approach may have been advantageous for the Pell grant recipients, based on a significant final exam outcome (p = 0.036) and large and medium effect sizes of g = 0.78 and g = 0.49 for the exam and CI, respectively. Similar results from the CI and multiple-choice portion of the final exam serve to triangulate the findings and strengthen the conclusion that the modified blended approach may be better in terms of cognitive-based outcomes, including those related to the lower levels of Bloom's taxonomy and conceptual understanding. For the free-response, open-ended final exam problems that targeted the higher levels of Bloom's taxonomy, the modified blended approach was not associated with significantly higher scores and only small effect sizes of $|g| < 0.20$.

Similar to the free-response problem outcomes, the affective outcomes from the CUCEI classroom environment inventory did not demonstrate the modified blended classroom to have the preferred classroom environment, despite students' enhanced performance on their CI and multiple-choice exam problems. This is based on non-significant differences and generally small effect sizes between the two approaches. This result may be related to students' anxiety with cumulative exams.

Acknowledgments – This research was made possible by the Engineering Education Research Center (EERC) in the Swanson School of Engineering at the University of Pittsburgh as well as the Research for Undergraduates Program in the College of Engineering at the University of South Florida. The work would not have been possible without the data from the control group collected under work supported partially by the National Science Foundation under Grant Number 1609637. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. R. Clark, A. Kaw and M. Besterfield-Sacre, Comparing the effectiveness of blended, semi-flipped, and flipped formats in an engineering numerical methods course, *Advances in Engineering Education*, **5**(3), 2016.
2. R. 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 and Learning*, **12**(1), Article 11, 2018.
3. R. Clark and A. Kaw, Benefits of adaptive lessons for pre-class preparation in a flipped numerical methods course, *International Journal of Mathematical Education in Science and Technology*, **51**(5), pp. 713–729, 2020.
4. R. Clark and A. Kaw, Adaptive learning in a numerical methods course for engineers: Evaluation in blended and flipped classrooms, *Computer Applications in Engineering Education*, **28**(1), pp. 62–79, 2020.
5. A. Kaw, G. Besterfield and J. Eison, Assessment of a web-enhanced course in numerical methods, *International Journal of Engineering Education*, **21**(4), pp. 712–722, 2005.
6. A. Kaw and M. Hess, Comparing effectiveness of instructional delivery modalities in an engineering course, *International Journal of Engineering Education*, **23**(3), pp. 508–516, 2007.
7. J. Bourne, D. Harris and F. Mayadas, Online engineering education: Learning anywhere, anytime, *Journal of Engineering Education*, **94**(1), pp. 131–146, 2005.
8. C. Dziuban, J. Hartman, F. Juge, P. Moskal and S. Sorg, Blended learning enters the mainstream, in C. Bonk and C. Graham (eds), *The Handbook of Blended Learning: Global Perspectives, Local designs*, John Wiley & Sons, Inc., San Francisco, CA, pp. 195–206, 2006.
9. D. Garrison and N. Vaughan, *Blended Learning in Higher Education: Framework, Principles, and Guidelines*, John Wiley & Sons, Inc., San Francisco, CA, pp. 4–8, 2008.
10. S. Lou, C. Chung, L. Chao, K. Tseng and R. Shih, Construction of a blended TRIZ creative learning platform, *International Journal of Engineering Education*, **28**(1), pp. 37–47, 2012.
11. R. Osguthorpe and C. Graham, Blended learning environments: Definitions and directions, *Quarterly Review of Distance Education*, **4**(3), pp. 227–233, 2003.
12. C. Twigg, Improving learning and reducing costs: Lessons learned from round I of the PEW grant program in course redesign, *Center for Academic Transformation, Rensselaer Polytechnic Institute, Troy, NY*, 2003.
13. D. Garrison and H. Kanuka, Blended learning: Uncovering its transformative potential in higher education, *The Internet and Higher Education*, **7**(2), pp. 95–105, 2004.
14. K. Henning, G. Bornefeld and S. Brall, Mechanical engineering at RWTH Aachen University: Professional curriculum development and teacher training, *European Journal of Engineering Education*, **32**(4), pp. 387–399, 2007.
15. A. Dollár and P. Steif, A web-based statics course used in an inverted classroom, *Proceedings of the ASEE Annual Conference and Exposition, Austin, TX, 2009.*
16. M. Restivo, J. Mendes, A. Lopes, C. Silva and F. Chouzal, A remote laboratory in engineering measurement, *IEEE Transactions on Industrial Electronics*, **56**(12), pp. 4836–4843, 2009.
17. J. Cortizo, E. Rodríguez, R. Vijande, J. Sierra and A. Noriega, Blended learning applied to the study of mechanical couplings in engineering, *Computers & Education*, **54**(4), pp. 1006–1019, 2010.
18. Z. Hu and S. Zhang, Blended/hybrid course design in active learning cloud at South Dakota State University, *2nd International Conference on Education Technology and Computer (ICETC)*, *IEEE*, 2010.
19. J. Méndez and E. González, A reactive blended learning proposal for an introductory control engineering course, *Computers & Education*, **54**(4), pp. 856–865, 2010.
20. R. Sell, S. Seiler and D. Ptasik, Embedded system and robotic education in a blended learning environment utilizing remote and virtual labs in the cloud, accompanied by 'robotic HomeLab kit, *International Journal of Emerging Technologies in Learning*, **7**(4), pp. 26–33, 2012.
21. T. Cavanagh, The blended learning toolkit: Improving student performance and retention, *Educause Quarterly Magazine*, **34***(*4), 2011.
22. J. Dunlosky, K. Rawson, E. Marsh, M. Nathan and D. Willingham, Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology, *Psychological Science in the Public Interest*, **14**(1), pp. 4–58, 2013.
23. P. Agarwal, J. Karpicke, S. Kang, H. Roediger III and K. McDermott, Examining the testing effect with open-and closed-book tests, *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, **22**(7), pp. 861–876, 2008.
24. H. Roediger III, A. Putnam and M. Smith, Ten benefits of testing and their applications to educational practice, in J. Mestre and B. Ross (eds), *Psychology of Learning and Motivation*, **55**, Academic Press, San Diego, CA, pp. 1–36, 2011.

25. R. Felder and R. Brent, *Teaching and Learning STEM: A Practical Guide*, Jossey-Bass, San Francisco, CA, pp. 116–117, 2016.
26. J. Dunlosky, Strengthening the student toolbox: Study strategies to boost Learning, *American Educator*, **37**(3), pp. 12–21, 2013.
27. H. Roediger III and J. Karpicke, Test-enhanced learning: Taking memory tests improves long-term retention, *Psychological Science*, **17**(3), pp. 249–255, 2006.
28. O. Adesope, D. Trevisan and N. Sundararajan, Rethinking the use of tests: A meta-analysis of practice testing, *Review of Educational Research*, **87**(3), pp. 659–701, 2017.
29. S. Sansgiry, A. Nadkarni and T. Lemke, Test Anxiety with respect to a comprehensive cumulative assessment, *Journal of Pharmacy Teaching*, **12**(1), pp. 41–59, 2005.
30. K. Szpunar, K. McDermott and H. Roediger, Expectation of a final cumulative test enhances long-term retention, *Memory & Cognition*, **35**(5), pp. 1007–1013, 2007.
31. M. Weimer, *Examining the Benefits of Cumulative Tests and Finals*, Retrieved from https://www.facultyfocus.com/articles/educational-assessment/examining-the-benefits-of-cumulative-tests-and-finals/, November 13, 2015.
32. N. Lawrence, Cumulative exams in the introductory psychology course, *Teaching of Psychology*, **40**(1), pp. 15–19, 2013.
33. M. Khanna, A. Brack and L. Finken, Short-and long-term effects of cumulative finals on student learning, *Teaching of Psychology*, **40**(3), pp. 175–182, 2013.
34. B. Fraser and D. Treagust, Validity and use of an instrument for assessing classroom psychosocial environment in higher education, *Higher Education*, **15**, pp. 37–57, 1986.
35. M. Norusis, *SPSS 14.0 Statistical Procedures Companion*, Prentice Hall, Upper Saddle River, NJ, pp. 545–549, 563–567, 2005.
36. A. Kaw, Y. Lou, A. Scott and R. Miller, Building a concept inventory for numerical methods: A chronology, *Proceedings of the ASEE Annual Conference and Exposition*, *New Orleans, LA*, 2016.
37. D. Quade, Rank analysis of covariance, *Journal of the American Statistical Association*, **62**(320), pp. 1187–1200, 1967.
38. A. Lawson, Rank analysis of covariance: alternative approaches, *The Statistician*, **32**(3), pp. 331–337, 1983.
39. J. Durlak, How to select, calculate, and interpret effect sizes, *Journal of Pediatric Psychology*, **34**(9), pp. 917–928, 2009.
40. J. Kotrlik, H. Williams and M. Jabor, Reporting and interpreting effect size in quantitative agricultural education research, *Journal of Agricultural Education*, **52**(1), pp. 132–142, 2011.
41. G. Sullivan and R. Feinn, Using effect size-Or why the p value is not enough, *Journal of Graduate Medical Education*, **4**(3), pp. 279–282, 2012.
42. T. Perneger, What's wrong with Bonferroni adjustments, *BMJ*, **316**, pp. 1236–1238, 1998.
43. A. Field, *Discovering Statistics using SPSS*, SAGE Publications, London, pp. 571–618, 2005.

Renee M. 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 at the University of Pittsburgh. She received the PhD degree in industrial engineering from the University of Pittsburgh. Dr. Clark focuses on assessment of research related to student and instructor professional development as well as active and experiential learning. Her research has been funded by the National Science Foundation as well as the University of Pittsburgh Office of the Provost.

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 from the Council for Advancement and Support of Education and Carnegie Foundation for Advancement of Teaching. Professor Kaw's main scholarly interests are in education research methods, open courseware development, flipped and personalized learning, bascule bridge design, fracture mechanics, composite materials, 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 open courseware in Numerical Methods (http://nm.MathForCollege.com). This courseware annually receives 1,000,000+page views (http://mathforcollege.com), 2,000,000+ views of the YouTube lectures (http://youtube.com/numericalmethodsguy), and 90,000+ visitors to the ''numerical methods guy'' blog (http://AutarKaw.org). This body of work is also used in the understanding of the impact of the flipped, blended, and adaptive settings on cognitive and affective learning gains of engineering students. Professor Kaw 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, Florida Senate Resolution, ASEE Prism, Times of India, NSF Discovery, and Voice of America.