Improving Student Learning Outcomes in Online Courses: An Investigation Into the Effects of Multiple Teaching Modalities

Julio César Bahamón*
Department of Computer Science
UNC Charlotte
jbahamon@uncc.edu

Audrey Rorrer†
Department of Computer Science
UNC Charlotte
audrey.rorrer@uncc.edu

ABSTRACT

We investigated learning outcomes within an undergraduate C Programming course that is taught in multiple modalities: in-person, online and blended. Because this course has been taught by the same instructor, using the same scaffolding activities, materials and approaches, we were uniquely positioned to conduct a quasi-experimental study of learning outcomes between courses and within students. The overarching goal was to glean knowledge and implications about assessment practices for undergraduate courses that are taught in multiple modalities. The objectives of our research are primarily to discern what differential impacts, if any, are found between the in-person and the online course delivery. We aimed to discover learning outcome patterns among the students who participate in these modalities. Findings from this study provide valuable information for undergraduate Computer Science programs by identifying any differential learning outcomes that students experience between in-person and online course instruction. The research questions addressed by the study were as follows: 1) What impact does modality have on student learning outcomes? 2) What patterns are discernable across student groups? 3) What relationship is there between final course grades and assignment module learning outcomes? In earlier work, we were surprised that no significant differences were obtained between course modality. While this was an encouraging finding, we believed that further data collection and analysis were needed, before making general conclusions about the two modalities. This paper reports on our efforts to collect additional data, while considering additional variables, such as instructor, multiple modalities, and online course design approaches.

CCS CONCEPTS

• **Social and professional topics → Computer science education**; *Student assessment*;

*Course Instructor.
†Assessment and Evaluation

SIGCSE '20, March 11–14, 2020, Portland, OR, USA

ACM ISBN 978-1-4503-6793-6/20/03…$15.00
https://doi.org/10.1145/3328778.3366880

KEYWORDS

course modality, online education, programming, retention, sex, gender, race

ACM Reference Format:
Julio César Bahamón and Audrey Rorrer. 2020. Improving Student Learning Outcomes in Online Courses: An Investigation Into the Effects of Multiple Teaching Modalities. In *The 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20), March 11–14, 2020, Portland, OR, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3328778.3366880

1 INTRODUCTION

In this paper, we report our results from an investigation of learning outcomes within an undergraduate *Introduction to C Programming* course that is taught in multiple modalities: fully face-to-face, hybrid (or blended) course and fully online. The course is taught in a face-to-face or hybrid modality during the Fall terms and in an fully online modality during the Summer terms. Because this course has been taught by the same instructor using the same scaffolding activities, materials and approaches, we were uniquely posited to conduct a quasi-experimental study of learning outcomes between courses and within students. The overarching project goal was to glean knowledge and implications about assessment practices for undergraduate courses that are taught in multiple modalities, with an emphasis on the effects of online instruction on student achievement. The project objectives were to discern what differential impacts, if any, are found between the face-to-face and the online modalities. In particular, we wanted to discover learning outcome patterns among the students who participate in these modalities. This study is descriptive and comparative, and it examines learning outcomes between course modality and between student groupings. Findings from the study of the different *Introduction to C Programming* course modalities provide valuable information for the undergraduate Computer Science program in terms of any differential degrees of learning that students experience between in person and online course instruction.

2 RELATED WORK

A considerable number of research studies have looked into the effectiveness of online Computer Science courses, particularly comparing the differences in student engagement and achievement between face-to-face and fully online or blended versions of the same courses. Results from these studies tend to indicate that student achievement is higher in traditional or flipped face-to-face courses than in online courses [1]. Nonetheless, it is important to consider that there are significant factors that may have a considerable impact on student success, which if effectively addressed could

reduce the achievement gap that seemingly exists. Some of these factors include but are not limited to course design and delivery aspects, such as instructor interaction, assignments and testing environments [8].

Dodero et al. compared two different Object Oriented Programming courses at two distinct universities. One course used a hybrid teaching modality and the other course used a fully online teaching modality [10]. The study focused on measuring class participation and overall success (pass/fail) rates. Results of their study indicated that the use of technologies and online tools was beneficial for improving student participation in the hybrid course during the face-to-face sessions; however, the same benefits were not observed in the online-only course. Nevertheless, there were no significant differences in the levels of student success between the two modalities.

McDonald et al. conducted a statistical comparison of course delivery modalities that looked at differences in student performance for Computer Networks and Database Systems courses [21]. Their analysis included an initial series of t-tests on student background characteristics, which were intended to rule out prior factors that could affect student performance. In addition to the t-tests, regression analysis was applied to the student demographic data to identify the key contributing factors. The study found that student's final course grades where significantly better for students in the face-to-face courses than for students in the online courses.

A study by Campbell et al. looked at the reasons for higher dropout rates in online CS1 courses, when compared with face-to-face flipped versions of the same course [8]. The authors identified a number of success factors that had a significant impact on student performance, such as the completion of practice exercises, practice exam problems in conditions similar to those in the course exams (both sections used paper-based, proctored tests), and interaction with the instructor.

One of the key challenges encountered by studies such as the ones listed above is the existence of previous factors or conditions that may influence student success in a college-level course (e.g. class level, prior academic performance and demographics) [21, 27]. This makes it necessary to use statistical analysis methods that take such factors into account if we are to obtain reliable results. To this effect, we employed the Propensity Score Matching technique in our study [27]. This methodology is described below.

2.1 Propensity Score Matching

Propensity Score Matching (PSM) is a methodology for comparing study participants, which creates treatment and control groups that are statistically equal [3, 4, 7, 12]. The PSM technique simplifies the process of matching study participants using multiple variables, by considering the probability that a participant in the treatment group with a given characteristic matches a participant in the control group. Typical uses of the PSM methodology have included assessment of student performance, such as studying whether first-year college seminars increase academic achievement and sense of belonging in undergraduate students [27, 29]. One of the principal reasons for using PSM is to control for selection bias in quasi-experimental studies; however, it is important to note that this technique does not eliminate self-selection bias in its entirety and is limited by the set of observable variables [23, 27]. An additional benefit of using PSM is that it can increase the interval validity of a study.

2.2 Uses of Propensity Score Matching in Higher Education

Dietrich and Lichtenberger used Propensity Score Matching in a study of the effects of beginning higher education at a community college and later transferring to a four-year institution [9, 29]. The study matched transfer students at a selective four-year college with *native* juniors at an equally selective four-year institution who graduated from the same high school. The study did not find any statistically significant difference between the two groups' bachelor degree completion rates.

3 PROJECT RATIONALE

The College of Computing and Informatics at UNC Charlotte is experiencing tremendous enrollment growth, with an increase of 77% in the last five years. Consequently, increasing instructional capacity is imperative, so that faculty are able to provide and sustain quality teaching and learning environments for our students. One of the methods being used to address the increase in enrollment is the offering of online courses, both as an alternative to face-to-face sections and as the only modality available in some cases.

With many of our courses being offered as entirely online courses, the questions we wish to address are: 1) What impact does modality have on student learning outcomes? 2) What patterns are discernable across student groups, i.e., do student traits help predict learning outcomes (e.g. native freshmen/transfer, cumulative credit hours, underrepresentation in computing)? and 3) What relationship is there between final course grades and assignment module learning outcomes? Findings from this study will inform our assessment practices across the undergraduate curriculum by signaling whether or not an online course modality produces comparable learning outcomes, and to what degree the learning outcomes in this course were achieved equitably across student groups.

4 COURSE DESCRIPTION

The *Introduction to C Programming* course is offered as a face-to-face (in person) or hybrid (blended) course during Fall terms, and as an online course during Summer terms. The course focuses on the study of the C programming language. Topics covered include data types, operators, functions, program structure, file I/O, storage classes, exceptions, concurrent programming, dynamic memory allocation and the C preprocessor. The course incorporates the use of active learning methodologies, which have been shown to improve student performance in undergraduate science, technology, engineering and mathematics (STEM) courses [6, 11]. A typical course module consists of lectures interspersed with exercises. The majority of the exercises consist of a small program, code fragment, or short exercise that is directly related to the topic covered in the current module. The exercises provide students with the opportunity to explore recently covered materials individually or with peers. Students submit the exercises at the end of class (or week in the online course), so the instructor can assess the class' comprehension of materials in a timely manner. Exercise submissions

are lightly evaluated by the teaching staff, i.e., the emphasis is on participation, student engagement and effort instead of correctness. Students in the face-to-face section work in small groups to complete these exercises. Previous work has shown the benefits of small learning groups [22]. The groups are structured using a simplified version of the Lightweight Teams approach developed by Latulipe et al. [17]. The students in the online section complete the exercises individually, without help from their peers. However, they can ask the course teaching staff for guidance.

The course also includes homework assignments that are completed outside of class. Homework consists of larger and more complex programming assignments that require students to incorporate several of the topics covered during the term and are designed to require several hours of focused effort. Homework is typically given one to two weeks for completion during the Fall term and five (5) to seven (7) days during the Summer term. It is important to note that Summer courses at our institution are either 5 or 10 weeks long. The Summer 2017 offering of the course was taught during a 5-week long session, whereas the Summer 2018 and Summer 2019 offerings of the course were taught during a 10-week long session.

The course also uses exams to evaluate students' knowledge of specific modules of information. Exam problems are similar in length to the exercises done in class. Some of the questions require students to write C code, without the use of a development environment or a C compiler. Exams during the Fall term are in a regular classroom and proctored by the course Instructor and Teaching Assistant. Exams during the Summer term are taken either remotely or at a testing facility; however, to preserve academic integrity, online exams are proctored by a third-party proctoring organization.

The course is primarily intended as an elective for Computer Science and Engineering students. Hence, it is expected that students already know at least one high-level programming language, such as Python or Java. Nevertheless, a considerable number of students have taken it as their first programming course. One of the notable differences is a higher percentage of Computer Science students in the Fall offerings of the course (49%) when compared to Summer course enrollment (24%). It is likely that this difference is the result of the course being part of the Distance Education program run by UNC Charlotte, which attracts students from across the state and from multiple majors during the Summer term.

4.1 Online Course Design

The online version of the course was created following the *Quality Matters* course design methodology [20]. Quality Matters is a set of standards intended to ensure that the course contents are presented to the students in a manner that is clear, well organized, and consistent. Additionally, Quality Matters promotes the notion that essential course components should be clearly linked to measurable learning outcomes and supported by appropriate educational materials. To this effect, the course uses a textbook, which is supplemented with PowerPoint slides, code samples and short video demonstrations prepared by the instructor. Students are also given links to instructional videos and tutorials provided by a third party. The same course design is utilized in the face-to-face, hybrid and online versions of the course.

The 'Quality Matters Approach' has been deployed in five (5) offerings of the same course: Summer and Fall 2017, Summer and Fall 2018 and Summer 2019. This approach requires course design to meet a very detailed set of requirements listed in the Quality Matters Rubric [19]. The rubric serves as a guideline to ensure that courses that are delivered either entirely or partially online are designed to incorporate best practices in online education. Essential to this approach is the idea of *Alignment*, which is intended to ensure that students achieve the course's learning outcomes. To this effect, the rubric provides parameters and recommendations for the implementation of what are considered essential course components [19]:

(1) Course Overview and Introduction
(2) Learning Objectives (Competencies)
(3) Assessment and Measurement
(4) Instructional Materials
(5) Learning Activities and Learner Interaction
(6) Course Technology
(7) Learner Support
(8) Accessibility and Usability

Both modalities of the course use the same materials, with the addition of face-to-face instruction and in-class active learning activities in the version of the course offered during the Fall term. In the Summer offering of the course, students complete all activities on their on, without the direct assistance of the instructor or Teaching Assistant. However, students have face-to-face access to the instructor during office hours and to both instructor and TA via the *Piazza* online forum platform [25].

4.2 Course Organization

We use the *Canvas* Learning Management System to administer the course content, assignments, tests, and grades [16]. Course content units are organized into modules. There is one module for each topic, its associated set of learning outcomes, and the assessment instruments that align with these. Figure 1 shows a sample of one of these modules in *Canvas*.

All course modules follow the same organization. Students are first presented with an *Activity Checklist*, which provides a listing of all the steps needed to complete the module. The checklist is followed by a high-level overview of the module's contents and list of topics. Next is a list of the course-level learning objectives covered by the module, including detailed learning outcomes specific to the module's topics. The following item in the module is the assigned reading(s) from the course textbook and/or external reference materials. Assigned readings are followed by *Prep Work Activities*, which are usually short quizzes or small exercises designed to verify that students have completed the reading assignment. These are followed by the the instructor's lecture slides. Some modules also include short videos or step-by-step demonstrations using small programs. Next is a set of one or more exercises for the students to practice the concepts covered in the module. These exercises usually ask the student to write a code fragment or provide a short answer to a question. Finally, each module has a set of *Review Activities*, which are conceptually similar to the *Prep Work Activities* and

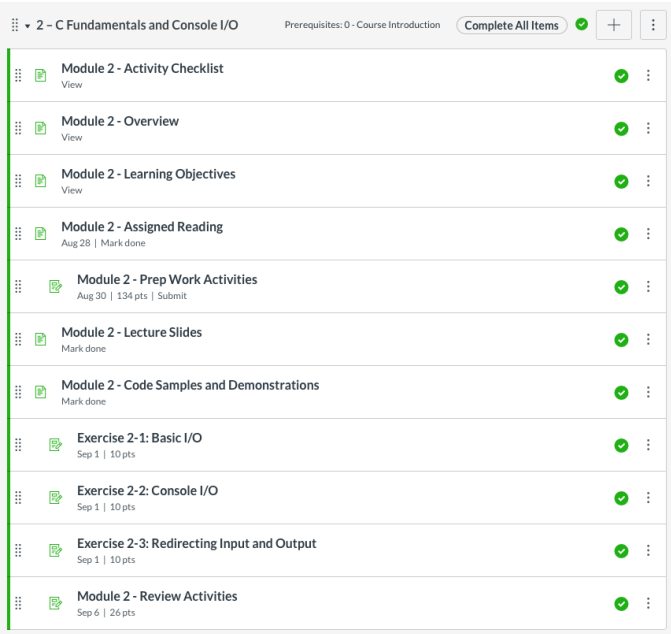

Figure 1: Sample Course Module in the *Canvas* LMS

to the Exercises, but designed to be significantly more challenging. Additionally, we use *Canvas* progress requirements to help ensure that students cover the module's components in the intended order. All modalities of the course use the same structure, with the exception that lecture slides, code samples and demonstrations, and exercises are presented by the course instructor during the lecture portion of the hybrid and face-to-face course. This also provides students with the opportunity to ask questions as the material is being discussed.

5 METHODOLOGY

The study design was a quasi-experimental 2x3 factorial design and includes multiple regression analysis. The research questions addressed were as follows.

(1) What impact does modality have on student learning outcomes?
(2) What patterns are discernable across student groups, i.e., do student traits predict learning outcomes (e.g. native freshmen/transfer, cumulative credit hours, underrepresentation in computing)?
(3) What relationship is there between final course grades and assignment module learning outcomes?

The sample size was small, therefore nonparametric statistical methods were employed to address the research questions. To address the first research question, is there a difference in learning outcomes by course modality and time, a Kruskal-Wallis H test was deployed, the nonparametric equivalent of an analysis of variance. The analysis included two course modalities: the Fall face-to-face modality, and the Summer online modality, and examined course outcomes of final grade and homework assignments. The grade distributions were not identical in the Fall and Summer courses, thus we compared the mean ranks. Additionally, we examined grade outcomes by High Performers, which were A and B final grades, and then of C grades, to determine if course modality resulted in different qualitative levels of performance.

A Chi-Square test of independence was used to examine the differences between student descriptive traits and learning outcomes, due to the small sample size not enabling multiple regression. The "underrepresentation in computing" variable was constructed by gender and ethnicity. Specifically, we created a "majority group" which included all Caucasian, Asian, and International status male students, and an "underrepresented group" which included all females, African American, Hispanic, Native American, Pacific Islanders, and Multi-Racial students.

Additionally, descriptive course data was examined, such as course evaluations. Course evaluations were distributed electronically and anonymously to students in both courses during the last week of the term.

Institutional Review Board approval was granted prior to the start of this study. Propensity score matching (PSM) was performed in *SPSS*, using the *R* plugin, to reduce the self-selection bias between Fall and Summer course enrollment based upon pre-course variables that are theoretically relevant [2, 23, 24, 26]. The PSM methodology utilized to configure the PSM algorithm follows the process recommended by Rocconi et al. [27]. Bipartite matching was performed for students based upon **level in school** at the time when they took the course (freshman, sophomore, junior, senior, graduate student), **overall grade point average** at the time they enrolled in the course , and **student origin type** (transferred into university or native freshman). Data pairs were created using a nearest neighbor matching, logistic regression and a caliper value of 0.2 [5, 13–15, 28]. We used the course modality as the treatment variable. The resultant matched sample was used to conduct comparisons between fall and summer course outcomes pertaining to the research questions. Figure 2 shows the demographic description among students in both courses.

6 PARTICIPANTS

In this section we describe the general statistics of the original sample, i.e., the data before PSM was applied, and the PSM sample, i.e., the set of matched sample pairs after PSM was applied.

6.1 Original Sample

There were a total of 86 students in the original sample, 61 from the face-to-face (fall) courses and 25 from the online (summer) courses. Gender and ethnicity distribution resembled the College's enrollment demographics: 88% were male students, 65% Caucasian, 3% Asian, 7% African American, 8% Hispanic/Latinx, 7% multi-ethnic/racial, and 5% international. The level in school break out was 45% junior level and 28% senior level; 41% were admitted to the university as native freshmen with 51% were having been admitted as transfer students. Forty-two percent were Computer Science majors.

6.2 PSM Sample

There were a total of 42 students in the PSM sample, 21 from the face-to-face (fall) courses and 21 from the online (summer) courses.

Total Number Enrolled	Fall Courses n=61		Summer Courses n=25		Propensity Score Matched n=42	
	number	percentage in course	number	percentage in course	number	percentage in sample
Student Level						
Freshmen	2	3%	0	0%	1	2%
Sophomore	7	11%	4	16%	2	5%
Junior	28	46%	11	44%	17	40%
Senior	17	28%	7	28%	17	40%
FifthYear & Special	7	11%	3	12%	5	12%
Gender						
Male	54	89%	22	88%	36	86%
Female	7	11%	3	12%	6	14%
Unreported	0	0%	0	0%	0	0%
Ethnicity						
African American	4	7%	2	8%	2	5%
Asian	2	3%	1	4%	1	2%
Caucasian	36	59%	20	80%	28	67%
Hispanic/Latinx	6	10%	1	4%	4	10%
Multi-racial/ethnic	6	10%	0	0%	2	5%
International	4	7%	0	0%	2	5%
Unreported	3	5%	1	4%	3	7%
Student Origin						
Native Freshman	29	48%	6	24%	12	29%
Transfer	29	48%	15	60%	28	67%
Unreported or Visitor	3	5%	4	16%	2	5%
Student Majors						
Computer Science	30	49%	6	24%	16	38%
Something Else	31	51%	19	76%	26	62%
Grades in Course						
A	22	36%	9	36%	13	31%
B	13	21%	5	20%	13	31%
C	8	13%	6	24%	9	21%
D	6	10%	3	12%	2	5%
F	12	20%	2	8%	5	12%

Figure 2: Demographic Data of Students Enrolled in Fall and Summer Courses.

Gender and ethnicity distribution resembled those in the original sample: 86% were male students, 67% Caucasian, 2% Asian, 5% African American, 9% Hispanic/Latinx, 5% multi-ethnic/racial, and 5% international. The level in school break out was 40% junior level and 40% senior level; 29% were admitted to the university as native freshmen with 67% were having been admitted as transfer students. Thirty-eight percent were Computer Science majors.

7 RESULTS

A Chi-square test was performed between all students in fall and summer courses to determine if grade distributions were different between gender (male, female) with no significant difference obtained. An additional Chi-Square test was performed to investigate whether or not grade distribution was different between majority race/ethnicity and minority race/ethnicity, with no significant difference obtained. These results suggest that gender and ethnicity were not factors related to outcomes for either course modality.

Results from the Kruskal-Wallis H test, performed on the subsample of 42 students derived from the propensity score matched group, indicated no statistically significant differences between Fall and Summer course outcomes for **course grades**, $\chi^2(2) = 0.027$, p = .870, mean rank of 21.81 for Fall and 21.19 for Summer. No significant differences were obtained for **homework grades**, $\chi^2(2) = .004$, p = .950, mean rank of 21.38 for Fall and 21.62 for Summer. When categorizing grades into **high performance**, meaning grades of *A* or *B* (n = 26), versus all else, no statistical significance was found between Fall (mean rank of 12.93) and Summer (mean rank 14.17), $\chi^2(2) = .169$, p = .681. When examining *C grades in the courses* (n = 9), no statistical difference was obtained; $\chi^2(2) = 2.94$, p = .086, Fall mean rank was 3.25 and Summer mean rank was 6.40.

Course evaluations were examined to determine what student perceptions were about the course. Fall courses evaluation had an average response rate of 54%. The majority (88%) of respondents agreed or strongly agreed with the statement that they learned a lot in the course, despite the fact that over half of the respondents were not Computer Science majors. Eighty eight percent felt the course challenged them to think, and 91% believed the course climate was conducive to learning. Open-ended comments as to what was helpful in the course from students indicated a positive relationship with the instructor, with the majority of the comments focusing on the instructor's knowledge and friendliness (15 out of 22 comments). There were a total of twenty three comments to the open ended question about improvements to the course. The theme from these comments was about pacing of the course; one student thought the class was too slow, while others thought it was too fast (n=3). Another theme was about the structure of the course in the Learning Management System; three students commented about logistical elements.

Summer course evaluation had a 58% response rate with 87% of respondents agreed or strongly agreed that they learned a lot in the course. 87% believed the course had challenged them to think. 80% thought the climate was conducive to learning. Seven students (47%) responded to the positive open ended comment box, both noting the quality of the resources and and course organization. Among the comments for areas of improvement, students noted the timing and length of modules and programming assignments to be of concern. One student commented in detail about the lack of availability for course discussion and interaction with the class for understanding materials.

The percentage of students with F grades was considerably higher in the face-to-face version of the course (20%) than in the online version (2%). Conversely, the percentage of withdrawals was higher in the online course (32%) than it was for the face-to-face course (22%).

8 DISCUSSION

Because the summer courses were condensed (5 or 10 weeks vs. 16 weeks), faster paced and completely online, we hypothesized that there would be different learning outcomes between course modalities. However, no statistically significant differences were found when comparing a matched subset of student grades. There is some indication from the course evaluations that students in the online (summer) modality struggled with material and pacing more so than in the face-to-face (fall) version of the course, albeit qualitative information. The theme from the face-to-face course centered around a positive perception of the instructor, versus the lack of the instructor in any comments by summer course students, which points to the importance of personalization of teaching to students. Several studies have noted the successful outcomes among students who feel engaged in a course, make friends, and connect to the professor [17, 18]. This factor is of particular interest if we consider that students in the fully online version of the course worked almost entirely on their own.

The different rates of failure and withdrawals between the two modalities are indicative of the nature of the student experiences. The lower failure rate, and higher withdrawals, in the online course may be indicative of students' early awareness of difficulty with the material, and without a personalized resource such as an instructor or peer group, students quickly decided to withdraw. Considering the limited amount of interaction of students in the online sections with the instructor or TA during office hours, it is possible that the effects of factors such as prior preparation, self-reliance or study habits become magnified and lead to earlier poor performance that in turn results in the early withdrawal that we recorded. Another possibility is that the shorter Summer term makes the decision whether to stay in a course more pressing, which coupled with the lack of a full-time enrollment requirement could explain the higher rate of withdrawals. It is also possible that the shorter time span of the Summer term does not give the instructor enough opportunity to make a significant impact on student performance.

9 CONCLUSIONS AND FUTURE WORK

In reviewing our analyses and findings that address the research questions, we were surprised that no significant differences appeared to exist between course modality. While this is an encouraging finding, we apply caution when considering other online modalities. The sample size and single course comparisons suggests we repeat this study across a wider variety of courses. We are planning future studies to account for additional variables, such as instructor, multiple modalities, and course design approaches other than *Quality Matters*. Future studies should be able to replicate these findings on a wider scale before we can generalize these outcomes.

The effects of different course design methodologies may be especially worthwhile to investigate. In particular, we would like to conduct a similar study on online courses that do not follow the *Quality Matters* standards in their design. Furthermore, comparing online with face-to-face courses that cover the same curriculum, but that are taught by different instructors or that use distinct design approaches may bring more clarity to one of the questions we set out to answer: whether course design approaches have a significant impact on the effectiveness of online courses? Based on the results of this study, we tend to believe that they do; however, this is a theory that needs to be tested.

Another question that has been raised as a result of this work is the effects of *self-selection bias* in online course performance. Specifically, we should study whether students' personal choice to register for the online version of a course rather than a face-to-face version introduces a bias in the student performance results recorded for such course. In other words, we need to consider the possibility that students who take online courses do so because they have prior personal experience that leads them to believe that they will perform well in such courses.

Finally, there are other aspects of online courses that this study hints at, but that we did not study. Some of these include the effects of a lack of personal interaction between instructor and student and the limited potential for collaboration with peers in an online setting. We believe that these factors deserve consideration, specially when we take into account the increased reliance on online courses to help manage enrollment growth.

ACKNOWLEDGMENTS

The authors would like to thank the UNC Charlotte Office of Assessment and Accreditation for their support of this research project. In particular, we want to recognize Dr. Elise Demeter for her guidance in the design of the study protocol.

REFERENCES

[1] 2018. Blended learning: Investigating the influence of engagement in multiple learning delivery modes on students' performance. *Telematics and Informatics* (2018). https://doi.org/10.1016/j.tele.2018.07.010

[2] Jill L. Adelson. 2013. Educational research with real-world data: Reducing selection bias with propensity scores. *Practical Assessment, Research & Evaluation* 18, 15 (2013), 2.

[3] Peter C. Austin. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research* 46, 3 (2011), 399–424.

[4] Haiyan Bai. 2011. Using propensity score analysis for making causal claims in research articles. *Educational Psychology Review* 23, 2 (2011), 273.

[5] Dimitri P. Bertsekas and Paul Tseng. 1988. Relaxation methods for minimum cost ordinary and generalized network flow problems. *Operations Research* 36, 1 (1988), 93–114.

[6] Charles C. Bonwell and James A. Eison. 1991. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports.* ERIC.

[7] Marco Caliendo and Sabine Kopeinig. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys* 22, 1 (2008), 31–72.

[8] Jennifer Campbell, Diane Horton, and Michelle Craig. 2016. Factors for Success in Online CS1. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '16).* ACM, New York, NY, USA, 320–325. https://doi.org/10.1145/2899415.2899457

[9] Cecile C. Dietrich and Eric J. Lichtenberger. 2015. Using propensity score matching to test the community college penalty assumption. *The Review of Higher Education* 38, 2 (2015), 193–219.

[10] Juan Manuel Dodero, Camino Fernández, and Daniel Sanz. 2003. An Experience on Students' Participation in Blended vs. Online Styles of Learning. *SIGCSE Bulletin* 35, 4 (Dec. 2003), 39–42. https://doi.org/10.1145/960492.960522

[11] Scott Freeman, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences* 111, 23 (2014), 8410–8415. https://doi.org/10.1073/pnas.1319030111

[12] Melissa M. Garrido, Amy S. Kelley, Julia Paris, Katherine Roza, Diane E. Meier, R. Sean Morrison, and Melissa D. Aldridge. 2014. Methods for constructing and assessing propensity scores. *Health services research* 49, 5 (2014), 1701–1720.

[13] Ben B Hansen. 2004. Full matching in an observational study of coaching for the SAT. *J. Amer. Statist. Assoc.* 99, 467 (2004), 609–618.

[14] Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis* 15, 3 (2007), 199–236.

[15] Daniel E. Ho, Kosuke Imai, Gary King, Elizabeth A Stuart, et al. 2011. MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, http://gking. harvard. edu/matchit* (2011).

[16] Instructure. 2018. Canvas Learning Management System. https://www.canvaslms.com/about-us/

[17] Celine Latulipe, N. Bruce Long, and Carlos E. Seminario. 2015. Structuring flipped classes with lightweight teams and gamification. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education.* ACM, 392–397.

[18] Mary Lou Maher, Celine Latulipe, Heather Lipford, and Audrey Rorrer. 2015. Flipped classroom strategies for CS education. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education.* ACM, 218–223.

[19] Quality Matters. 2018. Course Design Rubric Standards, 6th Edition. https://www.qualitymatters.org/qa-resources/rubric-standards/higher-ed-rubric

[20] Quality Matters. 2018. Standards from the Quality Matters Higher Education Rubric, 6th Edition. https://www.qualitymatters.org/sites/default/files/PDFs/StandardsfromtheQMHigherEducationRubric.pdf

[21] Merry McDonald, Brian Dorn, and Gary McDonald. 2004. A Statistical Analysis of Student Performance in Online Computer Science Courses. In *Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education (SIGCSE '04).* ACM, New York, NY, USA, 71–74. https://doi.org/10.1145/971300.971327

[22] Larry K. Michaelsen, Arletta Bauman Knight, and L. Dee Fink. 2002. *Team-based learning: A transformative use of small groups.* Greenwood publishing group.
[23] Antonio Olmos and Priyalatha Govindasamy. 2015. A Practical Guide for Using Propensity Score Weighting in R. *Practical Assessment, Research & Evaluation* 20 (2015).
[24] Antonio Olmos and Priyalatha Govindasamy. 2015. Propensity scores: a practical introduction using R. *Journal of MultiDisciplinary Evaluation* 11, 25 (2015), 68–88.
[25] Piazza Technologies. 2019. Piazza. https://piazza.com/product/overview
[26] R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. http://www.r-project.org
[27] Louis M. Rocconi, Jennifer Ann Morrow, and Sherry Marlow Ormsby. 2017. Utilizing Propensity Score Matching: A Practical Guide for Assessment Professionals. *Workshop at The Association for the Assessment of Learning in Higher Education (AALHE) conference* (2017).
[28] Felix Thoemmes. 2012. Propensity score matching in SPSS. *arXiv preprint arXiv:1201.6385* (2012).
[29] Angela L. Vaughan, Trent L. Lalonde, and Michael A. Jenkins-Guarnieri. 2014. Assessing student achievement in large-scale educational programs using hierarchical propensity scores. *Research in Higher Education* 55, 6 (2014), 564–580.