Toward CS1 at Scale: Building and Testing a MOOC-for-Credit Candidate

David Joyner
Georgia Institute of Technology, College of Computing
Atlanta, GA, USA
david.joyner@gatech.edu

ABSTRACT
If a MOOC is to qualify for equal credit as an existing on-campus offering, students must achieve comparable outcomes, both educational and attitudinal. We have built a MOOC for teaching CS1 with the intent of offering it for degree credit. To test its eligibility for credit, we delivered it as an online for-credit course for two semesters to 197 on-campus students who selected the online version rather than a traditional version. We compared the demographics, outcomes, and experiences of these students to the 715 students in the traditional version. We found the online students more likely to be older; to be underrepresented minorities; and to have previously failed a CS class. We then found that our online students attained comparable learning outcomes to students in the traditional section. Finally, we found that our online students perceived the online course quality more positively and required less time to achieve those comparable learning outcomes.

Author Keywords
Computing education, online learning, MOOCs for credit.

ACM Classification Keywords
• Social and professional topics~CS1 • Applied computing~ Computer-managed instruction

INTRODUCTION
The majority of Massive Open Online Courses (MOOCs) do not carry class credit [23]. This has allowed MOOCs to grow without the heavy burden of accreditation procedures; without degree credit, MOOCs typically do not need to give strong assertions about learning outcomes, academic integrity, or assessment validity.

However, as MOOCs have matured, interest has grown in using them for degree credit. Offering MOOCs for college credit may resolve overcrowded classrooms [18], lower the cost of college [29], reach audiences that otherwise cannot participate in these degrees [7], and improve pedagogy through the thoughtful use of new technologies [25]. However, critics note that online learning risks widening existing achievement gaps [16], diminishing student outcomes [28], lowering instructional quality [21], and losing valuable aspects of an on-campus experience [22].

In this work, we built and tested an Introduction to Computer Science (CS1) MOOC, then offered it for credit to on-campus students to test its credit worthiness. To maintain compatibility with the needs of MOOCs, all student work is automatically evaluated, exams are digitally proctored, and there are no mandatory components requiring human interaction. 215 students selected and completed the online course. Through pre- and post-tests and surveys, we compared the outcomes and attitudes of students in this version to those of the traditional version.

Earlier work has found conflicting results regarding whether MOOC-like courses like this one lead to the same learning outcomes as traditional courses [4, 6, 28]; in this work, we present the most comprehensive comparison yet of the learning outcomes of, attitudes about, and profiles drawn by different versions of classes like these. We are interested in answering three questions: (a) what types of students select the online version over the traditional version, (b) how do they perform compared to students in the traditional version, and (c) how do they rate the experience compared to the traditional version?

This work makes four main contributions to the existing literature. First, in line with our later hypothesis that the conflicting results from earlier research may be due to online course quality rather than student profiles, we present the design of this course for emulation by others. Second, we describe a profile of students that selected the online class. Third, we show there is no evidence that these online students learned less than their traditional counterparts. Fourth, we show there is evidence that these online students found their class experience preferable.

RELATED WORK
Early efforts towards developing MOOCs for credit were met with skepticism. One-time MOOC provider Udacity partnered with Colorado State University and San Jose State University to launch MOOCs for credit as far back as 2013. In the case of Colorado State University, few if any students applied to take the MOOC for credit [17]. At San Jose State University, students in traditional classes

L@S 2018, June 26–28, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5886-6/18/06…$15.00
https://doi.org/10.1145/3231644.3231665

outperformed students in a MOOC version [28]. These developments contributed to significant skepticism among faculty about whether MOOCs deserve course credit [11].

More recently, efforts toward for-credit MOOC-based education have gained more momentum. In 2014, Georgia Tech launched an online Master of Science in Computer Science program built around a collection of MOOCs. The courses themselves are not MOOCs, but rather use MOOCs in place of lectures while relying on typical admissions and registration processes and supplying customary human evaluation and office hours [12], mirroring a more blended model [10]. One study out of the program found comparable learning outcomes between online and on-campus students [9]. Coursera similarly offers four online Master's degrees, three with the University of Illinois and one with HEC Paris [19]. edX's MicroMaster's programs extend this approach further into the MOOC world, bypassing advanced admission into university programs and instead offering MOOCs that may be qualify for credit after admission [2].

At the undergraduate level, Arizona State University's Global Freshman Academy is the most prominent MOOCs-for-credit program [30]. It has been noted that very few of their enrollees are eligible to convert their MOOCs to credit [31], although research suggests completion rates may not be good metrics for evaluating MOOC success [27]. More recently, MIT and Georgia Tech have both begun experimenting with offering on-campus students access to MOOCs for credit with promising results [15, 20], and MIT found that a MOOC-based offerings led to comparable results as a traditional on-campus offering [4]. In this research, we build on this work by evaluating an offering of a MOOC-like CS1 class to on-campus students for credit.

SETTING THE STAGE

In this analysis, we will refer to three versions of CS101: CS101-Traditional, CS101-Online, and CS101-MOOC (sometimes shortened to Traditional, Online, and MOOC).

CS101-Traditional is the normal version of the course offered on-campus. Students attend three 50-minute 250-person lectures per week, and may attend one 75-minute 50-person recitation per week. Student grades mostly (70%) come from tests delivered on paper in person, but also from submitted homework assignments (30%). Students have access to a help desk for teaching assistant-run office hours.

CS101-Online is the experimental online version of the same course. The learning goals of the course are the same, but all components of the class are delivered online, including pre-filmed lectures, AI-graded homework assignments, an adaptive textbook, and digitally proctored exams. CS101-Online students have access to their own optional recitation as well as the help desk, but there are no required in-person or remote synchronous activities. Student grades are split evenly between tests (50%) and assignments (50%). There are mid-semester due dates, and institute requirements dictate that the course run along a normal 17-week semester. CS101-Online has been offered to on-campus students, but also through the institute's summer online undergraduate program and dual enrollment program for remote high school students. All students earn the same CS101 credit through CS101-Online as they would for CS101-Traditional.

CS101-MOOC will be the same course material as CS101-Online, but differ in student body and schedule. Whereas CS101-Online students must be enrolled at the institution, CS101-MOOC will be available to anyone. All lectures, tests, and assignments will be the same as or comparable to CS101-Online, but CS101-MOOC will be self-paced.

CS101-Online was designed to be compatible with CS101-MOOC: all assignments are automatically evaluated, all lecture material is recorded, and the tests are graded and proctored scalably. In fact, CS101-MOOC has already been as a non-credit MOOC, drawing 69,022 registrants, 12,422 active students, 519 verified certificates, and 665 course completers. We describe CS101-MOOC in the future tense to emphasize that our main interest is in the MOOC-for-credit; this existing MOOC course design's scalability.

To attach credit to CS101-MOOC, we must verify that students taking the alternate version are learning as much as students taking the traditional version. Here, we perform a pseudo-experiment to test CS101-Online against CS101-Traditional. We use surveys, pre-tests, and post-tests to examine students' experiences and learning outcomes in the context of potential demographic differences between the sections. We argue that if students enrolled at the university perform comparably in CS101-Online as students do in CS101-Traditional, and if the assessments and instruction in CS101-MOOC are comparable to those in CS101-Online, then CS101-MOOC is credit-worthy.

COURSE DESIGN

In describing CS101-Online, we differentiate development and delivery phases. Course development refers to components of the course that are produced before the semester begins and ideally may be reused; for CS101-Online, these are the videos, the textbook, and the assessments. Course delivery refers to those processes that cannot be performed in advance, like answering student questions and checking for academic integrity violations.

Lecture Videos

CS101-Online is made of 455 lecture videos averaging 2 minutes in length. These videos are divided into 5 units and 19 chapters. Approximately 10% of this video shows the instructor on screen in a studio introducing course principles alongside illustrative graphics, while 90% is screen capture of code demonstrations. During screen capture, the instructor uses a pen to write on top of the code and output while explaining its function. A screenshot from one such lecture video is shown in Figure 1. All code shown in the videos is also provided to students in a "sandbox" at the end of the lesson to allow them to immediately experiment with the code they have just seen.

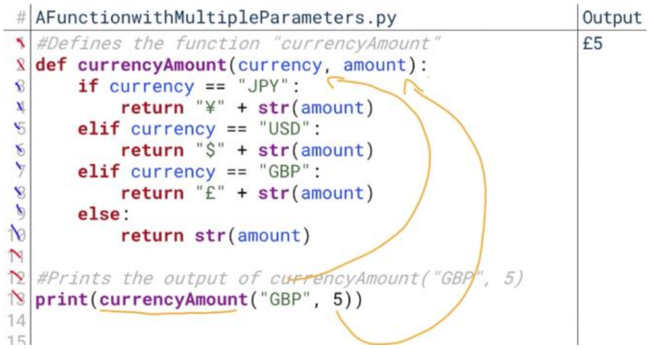

Figure 1. A screenshot from one of the course's lecture videos. Orange lines are drawn live during the video to demonstrate program function. Line numbers are crossed off in red when a line runs or blue when a line is skipped.

Textbook

A custom adaptive textbook was written for the class using McGraw-Hill's SmartBook platform. This textbook follows the same organization as the lectures. Visuals, sections, subsections, and examples are all shared between the two. This construction follows the principle of congruency [13] to allow students to switch between the videos and the textbook depending on personal preferences, environmental constraints, and the affordances of the respective styles. Students have reported that videos are preferable for initial learning, but the book is better supports study behaviors.

The textbook is interactive, supplying approximately 1,000 multiple choice and fill-in-the-blank problems for students to answer. These are required for course credit, making up approximately 5% of students' grades. Students will not answer all 1,000; instead, the book adapts question selection to students' demonstrated level of knowledge.

Course Assessments

There are two forms and three types of course assessments. The two forms are programming assessments, where students write and run code, and what we call 'small' assessments, which are fill-in-the-blank or multiple choice. The three types of assessment are exercise sets, problem sets, and tests:

- Exercise Sets: assessments interspersed between lecture videos, typically at least one after every video. Smartbook exercises are included in exercise sets.
- Problem Sets: Groups of 10-15 slightly more challenging assessments presented after the chapter.
- Tests: Timed, digitally proctored sets of 10-12 questions open for a limited time.

All assessment types include both assessment forms and run entirely in the browser. Small assessments in the exercise sets allow unlimited submissions, while small assessments in the problem sets and tests limit the number of submissions available. Programming assessments, including those on the tests, are presented in a simple in-browser development environment, which includes an output window for showing both program output and submission results. All programming assessments are autograded against a set of both pre-written and intelligently-generated test cases. This autograder shows students the complete results of their submissions: what inputs were used, what result was observed (including error information if applicable) from their code, and what result was expected. Upon completion, the autograder gives students one or more sample answers against which to compare their code to show alternate and optimal approaches, as well as good programming style. Students have unlimited attempts on every programming assessment to encourage revision; as a result, no partial credit is awarded on programming assessments. An example of a programming exercise and its autograder output are shown in Figure 2. A more complete explanation of the AI underlying these exercise evaluations can be found in [14].

Altogether, students complete approximately 250 programming assessments and 1,000 small assessments. At any time (including during tests), students may view their current course grade, which updates live as they complete assessments.

COURSE DELIVERY

Once the semester begins, students are given a 17-week calendar (12 weeks in summer) for course completion. Each major unit deadline coincides with that unit's test; the unit test typically occurred 1 to 2 weeks after the recommended deadline for the final chapter of that unit to allow on-pace students time to seek help before the test.

Course delivery involves five components: test proctoring, instructor communication, question-answering, recitations, and the help desk. These responsibilities are handled by the

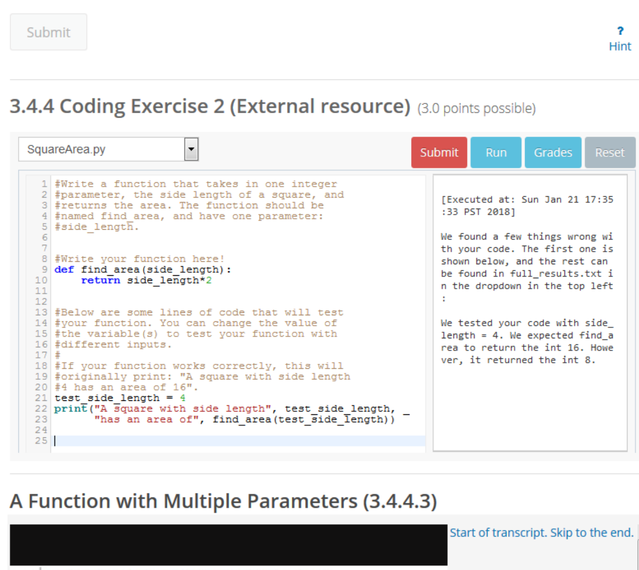

Figure 2. A programming exercise interleaved between a small assessment and a lecture video. On the left are the directions and student code; on the right is the live feedback from the autograder. The dropdown gives access to the full autograder output and sample answers after the exercise is completed.

Table 1. Enrollment in CS101-Traditional and CS101-Online since Spring 2017.

	Spr. 17	Sum. 17	Fall 17	Spr. 18
Traditional	386	32	329	326
Online	59	27	138	216

instructor and the teaching team. Recitations and the help desk are run according to the usual schedule process of the Traditional class; they are generally lightly-used by students in the Online section. We note that the other three components presently rely on human interaction, but that (a) work is underway—both by us and by others [26]—to create scalable methods for addressing these components; (b) certificate costs associated with for-credit MOOCs, while low, are sufficient to offer this human interaction to MOOC-for-credit students if necessary; and (c) to maintain credit-worthiness, we must assert what students know exiting the course, not the support they needed to get there.

Test Proctoring

Students complete four closed-book proctored tests during the semester, one for each major unit. Tests are typically open for four days, given research indicating that such exams are not more prone to cheating behaviors [3].

Tests are digitally proctored by a service that uses students' webcams, microphones, and screen capture to digitally record the test session. This data is then automatically analyzed, and possible instances of forbidden behaviors—such as excessively looking away from the screen, the appearance of another face or voice, or accessing forbidden web sites—are flagged. Due to institutional requirements, the teaching team then reviews these flags; however, for CS101-MOOC students, the proctoring service may review flags. Thus, we call this proctoring process "asynchronous streamlined" proctoring: no student is automatically penalized for suspected collaboration, but rather humans rapidly review flags after the test is completed.

Instructor Communication

The instructor in the Online version engages in two primary forms of communication: announcements and individual emails. Every week the instructor posts an announcement reminding students of upcoming recommended deadlines, imminent tests or surveys, and anything else they need to know for that week. This emphasizes pushing information to students rather than requiring them to pull information at appropriate times, mimicking the routine that comes from attending class at prescheduled times.

Secondly, once a week, the instructor individually emails any student who has fallen one or more weeks behind the recommended schedule, in part to help online students feel less isolated. It has been noted that this is not inherent to the online nature of the course; however, we note that this is an instance of reinvesting into course delivery the time saved by pre-producing lecture material.

Question-Answering

Lastly, students may ask questions of the teaching team. Typically, these are debugging questions about their code, in addition to some questions on principles and CS as a whole. Students are offered a course forum (Piazza) and a course chat room (Slack) for these questions. The instructor monitors both and answers most questions; in Fall 2017, students in the Online section reported that the instructor answered their questions 75% of the time and their classmates answered 25% of the time; students in the Traditional version reported that TAs answered their questions 55% of the time and classmates 45% of the time. Again, we note this increased instructor attentiveness is due to reinvesting the time saved by pre-producing material.

METHODOLOGY

To test the course, we conducted a pseudo-experiment comparing CS101-Online to CS101-Traditional. Both courses have been available for university students since Spring 2017, and students may select which version to take. Table 1 shows the enrollment patterns in each version during this time. In this analysis, we focus on Spring 2017 and Fall 2017.

Table 2. Course and survey completion data. Tr. Stands for Traditional, On. stands for Online. Pre-Course Survey and Pre-Test percentages are of initial enrollment; Post-Course Survey, Post-Test, and Set percentages are of completion numbers. Pre-Test & Post-Test Set counts the individual students that completed both the pre-test and the post-test. In Fall 2017, students were asked at the end to report if they put in their best effort; High Effort reflects these numbers.

	Spring 2017		Fall 2017	
	Tr.	On.	Tr.	On.
Enrollment	386	59	329	138
Course Completion	332	58	290	133
Withdrawals	54 14%	1 2%	39 12%	5 4%
Pre-Course Survey Response	328 85%	55 93%	272 83%	127 92%
Post-Course Survey Response	262 79%	53 91%	158 54%	120 90%
Pre-Test Response	269 70%	55 93%	248 75%	123 89%
Post-Test Response	227 68%	46 79%	115 40%	122 92%
Pre- & Post-Test Set	193 58%	45 77%	89 31%	115 86%
Pre-Test High-Effort Response	*	*	153 47%	82 59%
Post-Test High-Effort Response	*	*	55 19%	50 38%
Pre- & Post-Test High Effort	*	*	40 14%	37 28%

Due to FERPA and IRB constraints, only data that can be exported without personally identifying information is included; however, most tools used in this analysis export data anonymously. Additionally, it is possible that students in the online version attended the traditional class, or that students in the traditional class accessed the MOOC version of the online material. A survey question has been added to future semesters to measure whether this occurs.

RESULTS

To present the results, Table 2 first presents the rates of course completion, survey completion, and test completion across the sections and semesters. It is worth acknowledging that the online section has reliably had higher response rates. As both sections attribute approximately incentives for survey and test completion, it is unclear if this is due to systematic differences in the conscientiousness of the online student body or differing levels of encouragement to complete the surveys.

In reporting the results, we focus on three components: demographic data, performance data, and attitudinal data.

Demographic Data

Demographic analysis serves two purposes. First, the findings are noteworthy on their own: it is interesting that the two sections draw notably different demographics of students. Second, it allows investigation into whether differences in the sections are due to class structures or student differences. As this is a pseudo-experiment without random assignment, our conclusions are inherently limited by differences in the student bodies between the sections. Table 3 gives the demographic comparisons between these versions across both semesters.

These demographic data highlight several notable differences between students that select each section. First and most notably, the differences changed considerably between Spring 2017 and Fall 2017; we hypothesize this is partially due to systematic differences in when different students that take the class, and partially due to a registration cap in Spring 2017 that rewarded students who were more attentive to registration. Fall 2017 offered more free selection, and so we focus on this in our analysis.

In Fall 2017, students in the online class were more likely to be older than 20 ($X^2 = 5.037$, $df = 1$, $p < 0.05$), to be 3rd year or later in school ($X^2 = 7.893$, $df = 1$, $p < 0.01$), and to identify as an underrepresented (non-White or Asian) minority ($X^2 = 4.137$, $df = 1$, $p < 0.05$). The distribution of majors differed significantly ($X^2 = 28.741$, $df = 5$, $p < 0.001$), with the Online section preferred by Business and Engineering majors while the Traditional section was preferred by Computing, Science, and Math majors. Online students were more likely to be employed ($X^2 = 13.919$, $df = 1$, $p < 0.001$). Prior experience differed between the sections ($X^2 = 21.348$, $df = 5$, $p < 0.001$), but no difference exists in prior expertise, whether defined to include only prior completion of a computer science course ($X^2 = 0.035$, $df = 1$, $p = 0.8525$) or self-taught programmers as well ($X^2 = 0.622$, $df = 1$, $p = 0.4302$).

Table 3. Demographic differences in students that elect to enroll in the traditional and online sections. All numbers are percentages; due to rounding or multi-selection, numbers may not add to 100. Asterisks indicate questions or options that were not available in that semester's survey.

	Spring 2017		Fall 2017	
	Tr.	On.	Tr.	On.
What is your gender?				
Female	54	49	45	46
Male	46	49	55	54
Other	0	2	1	0
No Answer	0	0	0	0
How old are you?				
<18	0	0	2	1
18-20	73	47	88	81
21-23	10	34	9	12
24+	2	14	1	6
What year are you?				
1st year	*	*	72	60
2nd year	*	*	18	18
3rd+ year	*	*	11	22
What is your race/ethnicity?				
White	61	57	58	60
Black or African-American	5	13	9	13
Hispanic, Latino, Spanish origin	6	4	3	8
Asian	24	23	29	33
Other	5	4	2	1
Decline to Answer	*	*	2	1
How much prior programming experience do you have?				
Took an in-person course	20	31	*	*
Completed an in-person course	*	*	26	26
Failed/quit an in-person course	*	*	3	10
Took an online course	6	7	*	*
Completed an online course	*	*	6	10
Failed/quit an online course	*	*	4	2
Self-taught	9	20	17	14
No prior experience	65	49	44	38
What is your employment status during the semester?				
Employed full-time	*	*	1	4
Employed part-time	*	*	17	30
Not employed	*	*	82	65
What is your major?				
Business	*	*	6	16
Computing	*	*	38	24
Engineering	*	*	20	25
Liberal Arts	*	*	6	6
Science & Math	*	*	24	18
Other	*	*	6	11

In summary, our Online students were more likely to be older; to be underrepresented minorities; to be majoring in Engineering or Business; and to have previously failed or dropped out of a CS class. These demographic differences generally held between the two semesters. We hypothesize there are significant interaction effects between those trends; for example, research suggests underrepresented minorities are more likely to have to balance work and school [1]. Future work will focus on the reasons for and interrelationships between these demographic trends. We hypothesize these trends are in part because certain demographic groups need more flexibility in their coursework, and that perceptions of the rigor of different versions of the course alter student selections.

In addition to this demographic survey, in Fall 2017 we also asked students why they selected their version. Among students that selected the Online version, time flexibility was the top primary reason (33%), followed by scheduling conflicts (22%) and a preference for self-paced learning (18%). Among those that selected the Traditional version, a preference for live lecture meetings was the top primary reason (46%), followed by a preference for structured class meeting times (28%). 29% of students in the Traditional section reported they were unaware of the Online class.

Performance Data

During the first and last weeks of the semester, students in both sections take the Secondary CS1 (SCS1) knowledge assessment [24], a 27-item evaluation of CS1 knowledge. Students receive class credit for completing it, but their score on the assessment is not weighed into their average. In evaluating performance data, we examine two measurements: outcomes (post-test scores) and change (post-test scores minus pre-test scores).

Table 4 gives the pre-test, post-test, and change scores for students in both sections of both semesters. In Fall 2017, we added a question to the end of the tests asking students to report whether they put in their best effort on the test; we noted this would not affect their score, but it would influence our analysis. For Fall 2017, we include specific results for those students who reported high effort.

Altogether, there is no evidence that students in the Online section are disadvantaged compared to the students in the Traditional section. Students in the Online section that reported high effort on the post-test scored better than similar students in the Traditional section on the Fall 2017 post-test with statistical significance. The apparent contradiction with the lack of a statistically significant change in for these groups arises from the 28 students who reported high effort on the post-test, but not on the pre-test; their data is excluded from the Change value.

Overall, we conclude that there is no evidence Online students perform worse than Traditional students. There is some evidence for superior performance in the Online section, but more systematic and consistent evidence to that end would be necessary to conclude Online students achieve greater knowledge.

Table 4. Average Pre-Test, Post-Test, and Change scores in both sections and semesters. Standard deviations are given in parentheses; statistical tests of equality are given below values. For Change, only students that completed both the Pre-Test and Post-Test are included. See Table 2 for *n* values.

		Spring 2017		Fall 2017	
		Tr.	On.	Tr.	On.
Pre-Test		6.78 (4.06)	7.36 (3.30)	7.22 (4.04)	7.47 (4.43)
		$t = 1.14$, $p = 0.26$		$t = 0.54$, $p = 0.59$	
	High Effort			7.93 (4.62)	8.28 (4.76)
				$t = 0.55$, $p = 0.58$	
Post-Test		9.73 (4.39)	10.78 (4.66)	10.50 (5.01)	11.29 (4.97)
		$t = 1.47$, $p = 0.14$		$t = 1.21$, $p = 0.23$	
	High Effort			12.29 (5.36)	14.84 (4.87)
				$t = 2.54$, $p < 0.01$	
Change		+2.91 (4.66)	+3.07 (4.77)	+4.22 (4.54)	+3.66 (5.03)
		$t = 0.21$, $p = 0.84$		$t = 0.81$, $p = 0.42$	
	High Effort			5.50 (4.81)	5.46 (4.19)
				$t = 0.04$, $p = 0.97$	

Attitudinal Data

Merely equaling the same learning outcomes does not necessarily guarantee that the Online course ought to be offered for credit; if, for example, students in the Online course come away with a significantly more negative impression of the course or CS as a whole, it would still be ill-suited to broadcasting to a larger audience for credit.

So, we also evaluate the attitudes of students exiting each version. Many of these questions are evaluated using 7-point Likert scales; for these questions, we summarize results with an interpolated median, but statistical significance was calculated with a two-tailed Mann-Whitney *U* Test. Other questions are nominal; for these, we show the entire distribution and use a Chi-square test.

Overall Perceptions

We are first interested in overall perceptions of the courses. Toward this end, the post-course survey asks several questions regarding the classes' pace, rigor, quality, and relative value compared to other college courses. On all questions, 1 represented a strongly negative value ("Way Too Fast", "Way Too Hard", "Not Nearly as Good", "Bad"), 7 a corresponding strongly positive value, and 4 a neutral value. Table 5 shows these comparisons.

Table 5. Student perceptions of their course's pace, difficulty, and quality both on its own and as compared to other courses. For Pace and Rigor, 4 represents "About Right", while higher and lower represent too fast/hard or too slow/easy. For Quality, higher scores represent higher perception of quality.

	Spring 2017		Fall 2017	
	Tr.	On.	Tr.	On.
Course Pace	4.30	4.20	4.23	4.07
	$Z = 0.66, p = 0.51$		$Z = 2.18, p = 0.03$	
Course Difficulty	4.63	4.54	4.62	4.11
	$Z = 0.23, p = 0.81$		$Z = 4.27, p < 0.01$	
Quality vs. Other Courses	4.93	5.96	5.37	6.07
	$Z = -4.61, p < 0.01$		$Z = -4.61, p < 0.01$	
Overall Quality	5.20	6.13	5.58	6.35
	$Z = -4.70, p < 0.01$		$Z = -5.09, p < 0.01$	

In Spring 2017, students in both courses perceived their version as slightly too fast and slightly too hard. Students in the Online section, however, had a significantly more positive view on the course. In Fall 2017, students in the Online version still perceived their version as higher-quality, but also perceived its pace and rigor as more appropriate. We hypothesize this is due to changes made between the semesters that allocated more calendar time to challenging topics and less to easier topics. We also note that the flexibility to easily adjust the schedule in this way is due to the online, pre-produced nature of the course.

Given the equal results on performance measures we reflect positively on students' perceptions of pace and rigor moving closer to the midpoint of the scale. Computer science is often regarded as a hard, unwelcoming topic [8], and we embrace students perceiving it as more manageable so long as the learning outcomes remain comparable.

Specific Components

Second, we are interested in perceptions of the value of individual pieces of the different classes. We identified six components that are used in both versions. For each component, we asked students to agree or disagree on a 7-point Likert scale with the statement: "The [component] was valuable in helping me learn the material." Table 6 presents our results for each of these components.

In Spring 2017, three statistically significant differences were observed: Online students valued the Lectures, Assignments, and Forums more highly than Traditional students. In Fall 2017, those three differences were replicated, and two more differences were observed. Traditional students valued Recitations more, while Online students valued Tests more. This new observation regarding Recitation value likely comes from the increased sample size in Fall 2017; however, we hypothesize that the increased value attached to tests is due to improvements made to the Online course tests between the semesters.

Table 6. Student perceptions of the value of each of six components common to both the Online and Traditional versions. Numbers shown are interpolated medians.

	Spring 2017		Fall 2017	
	Tr.	On.	Tr.	On.
Lectures	5.34	6.34	5.73	6.60
	$Z = -4.50, p < 0.01$		$Z = -4.74, p < 0.01$	
Recitation	4.84	4.11	5.16	4.03
	$Z = 1.90, p = 0.06$		$Z = 4.62, p < 0.01$	
Textbook	4.95	4.22	4.43	4.10
	$Z = 1.07, p = 0.28$		$Z = 0.60, p = 0.55$	
Assignments	6.26	6.73	6.36	6.75
	$Z = -2.16, p < 0.05$		$Z = -2.90, p < 0.01$	
Tests	5.04	5.12	5.39	6.08
	$Z = -0.07, p = 0.94$		$Z = -3.88, p < 0.01$	
Forum	5.24	6.03	5.27	6.00
	$Z = -3.57, p < 0.01$		$Z = -2.49, p < 0.01$	

It is not clear to what extent these differences are inherent between online and traditional classes and to what extent they are attributed to these specific implementations; however, we argue that these implementations aim to take advantage of opportunities in teaching online at scale. Although these differences may not be automatic when transitioning to online at scale, they are uniquely possible.

Personal Impressions

Third, we are interested in some elements of the students' personal experience with the course. For example, one common critique of online learning is that students feel isolated and are unable to find help when needed, or do not develop the same confidence in their ability.

To evaluate this, we asked two relevant questions. First, we asked students to agree or disagree on a 7-point Likert scale with the statement, "I feel like I know where to seek help if I get stuck." Second, we asked students to rate their confidence in their class performance on a **5**-point Likert scale from "Very unconfident" to "Very confident". The results of these comparisons are presented in Table 7.

Table 7. Students' self-reported comfort seeking help and confidence in their class performance at the end of the semester. Numbers shown are interpolated medians. The first question was not offered in the Spring 2017 survey.

	Spring 2017		Fall 2017	
	Tr.	On.	Tr.	On.
Know Where to Seek Help	*	*	6.35	6.31
			$Z = 0.21, p = 0.83$	
Confidence in Performance	3.97	4.07	4.03	4.17
	$Z = 0.25, p = 0.80$		$Z = -1.34, p = 0.18$	

We see no evidence that students differed in their level of confidence or their ability to find help when needed.

Time on Task

Fourth, we are interested in the amount of time that students invest into each version of the course. One argument in favor of online learning is that it improves efficiency: in addition to removing sources of lost time (classroom technical issues, leftover lecture time, etc.), students can focus on their just-right level of difficulty rather than losing time while confused in a synchronous lecture.

To evaluate this, we asked students to estimate how much time they spent on the course, including all activities (attending or watching lectures, completing assignments, studying for and taking tests, etc.). The results of this comparison are shown in Table 8.

We observed a significant difference in reported time spent on the class, which became larger in the second semester. The most likely explanation for this would be that the Online version requires less work; to investigate this, we acquired the homework assignments from both the Online and Traditional sections in Fall 2017. We counted that Traditional students completed roughly 60 programming problems (11 assignments with ~5 problems each) which together counted for 25% of their averages. Online students completed approximately 250 programming problems and 1,000 multiple choice or fill-in-the-blank exercises, which together counted for 50% of their averages. The scope of the individual problems approximately matched between the sections, suggesting approximately equal workload.

Demographic differences may explain a portion of this difference, but given the significant change, we conclude that the Online version is demanding less time. Given the equivalent performance results, this suggests that students in the Online section are learning more efficiently, whether due to the flexibility, the instructional quality, or efficient use of time. Regardless, this presents an opportunity to build more coursework and practice into the Online section and to investigate how all students spend their time.

Attitudinal Summary

The attitudinal surveys give a positive picture of students' perceptions of the Online version: they rate it as higher quality overall, perceive its pace and rigor as slightly more appropriate, and have a greater appreciation of certain individual components. There is no evidence that they are uncertain about where to seek help, countering the idea that online students are isolated and that online learning should only be used by highly self-regulated learners. Interestingly, students in the Online version also report spending significantly less time per week on course material.

Table 8. Self-reported hours spent on coursework across the two semesters and versions of the course. Numbers shown are percentages of students responding to the survey.

	Spring 2017		Fall 2017	
	Tr.	On.	Tr.	On.
<5 hours per week	9	17	12	33
5 to 7 hours per week	32	33	31	49
8 to 10 hours per week	33	45	38	11
11 to 13 hours per week	19	3	13	5
14 or more hours per week	7	2	6	2
	$X^2 = 28.55$ $p < 0.01$		$X^2 = 73.98$ $p < 0.01$	

We attempted to evaluate attitudes and behaviors in other ways without success. First, we surveyed students regarding their likelihood to switch to CS as their major or add a CS minor, but observed no difference between the sections. Students noted that likelihood to switch majors or add a minor is a product of many factors beyond appreciation of the subject, and so it would be inaccurate to attribute changes to the course experience on its own.

We also found attendance of the help desk and recitations to be dramatically lower in the Online version than the Traditional version. 70% of Online students attended two or fewer recitations (45% attended none), while 89% of Traditional students attended 3 or more. 84% of Online students never visited the help desk, while 70% of Traditional students did at least once. These findings are notable because these components are among the only parts of CS101-Online that are not massively scalable. Their low usage suggests they are not responsible for the positive results associated with the Online class.

CONCLUSION

In this paper, we have compared the performance, behaviors, and attitudes of students in a traditional, face-to-face CS1 class to those of students in an online version of the same course material. We found that Online students learned as much as students in the Traditional version. We also found that students in the Online version perceived the course as higher quality and reported needing less time to reach those learning outcomes. They also more positively perceived the lectures, assignments, and course discussion forum. These results were first observed in a pilot experiment in Spring 2017, and were replicated with a large sample size in Fall 2017. During the replication experiment, additional differences emerged: Online students perceived the pace and rigor of the course as more appropriate and viewed the course tests as valuable learning experiences.

We thus conclude that a CS101-MOOC, built from the same instructional material and using the same assessments, would be credit-worthy. The learning outcomes derived from the Online course are comparable to those of the Traditional course and the student experience is positive.

Limitations & Future Work

There are many limitations to this study; future work ought to focus on addressing these limitations to find the degree to which these findings may be transferred.

Methodological and Statistical Limitations

First, we performed 42 statistical tests in this work, with 16 significant at $\alpha = 0.01$ and 3 more significant at $\alpha = 0.05$.

These correspond to family-wise error rates of 34% and 88% respectively, meaning that it is likely that there are Type I Errors, but it is unlikely ($p < 0.05$) that there are more than two errors among those significant at $\alpha = 0.01$.

Second, much of this study hinges on the accuracy of the language-neutral CS1 assessment. While validated on its own [24], students in both sections still end the course answering fewer than 50% of the questions correctly. This combined with low completion rates may cast doubt on whether the tests are accurately capturing learning outcomes. We plan to address this limitation by better incentivizing test completion in future semesters and tracking the outcomes of students in their future CS classes.

Last, there is a seeming contradiction that must be explored. Online students reached similar learning outcomes as Traditional students while reporting less time interacting with the course material, but they also completed more problems. If students completed more problems, how did they spend *less* time on the course? If this is because they solve problems more rapidly, then should that not be shown on the post-test? To examine this, we will investigate how students spend time with the courses.

Bias Limitations

If it is true that students emerge with comparable learning outcomes, we ask: what causes the other positive differences? Are they due to inherent differences between Online and Traditional, or to differences more specific to these two offerings? Investigation into causes of these positive differences must take place in other classes as well due to a biasing Halo effect [5] where students, liking the course overall for any reason, rate each component highly as a reflection of that overall perception.

Finally, in addressing one of our research questions, we bias another. Students self-select into their desired version of the course, and so we can research who selects which version and why, but we cannot attribute differences in outcomes strongly to version differences rather than student differences. We measure this confound with pre-tests ensuring that incoming student ability is not significantly different across the versions, but other elements of their backgrounds may bias the results in other ways. To truly measure the effectiveness of the online material, a controlled experiment would have to be conducted.

Generalization Limitations

While there is value in creating an online version of this class on its own, one of the stated goals of this work is to create a MOOC-for-credit. However, there are portions of the on-campus experience that still would not translate to a MOOC, including the presence of a help desk and recitations, the social interactions among collocated peers, and the admissions process that acts as a filter for which students may take the course. In offering the MOOC for credit, we would argue that our emphasis ought to be on ensuring that a student cannot receive credit without reaching the same learning outcomes. Where they receive support and whether they are likely to succeed are interesting questions, but as long as the assessments remain valid measures of the learning goals, the route taken does not threaten the viability of offering credit. Monitoring and testing, however, would be necessary to ensure that a disconnect does not arise between the assessments and the learning goals as MOOC students earn credit.

Confounding Limitations

Generalizing out from this work to the ongoing investigation of learning outcomes in online courses and MOOCs, it is important to address a confound in the broader literature. In this and prior work that showed equivalence between MOOC-like offerings and traditional offerings [4], students were drawn from the on-campus student body of a major public research university, and content was produced by experts from that university. In work finding that online students underperform, the student body is drawn from more open community colleges [28], and the content is produced and delivered by those more resource-constrained colleges or young, unproven start-ups. It is possible that the proficient, privileged, or self-regulated students admitted to top-tier colleges are able to succeed online, and students who gravitate toward community colleges need traditional classes [6]. It is also possible that these large universities are better equipped than smaller colleges or start-ups to create effective online courses. To investigate this, work is already underway to open access to this CS1 course to other colleges, as well as to use as a basis for high school CS classes in underprivileged schools.

Future Work

Thus, future work—both our own and the community's—ought to focus on further examining the credit-worthiness of MOOC-style courses, including measuring their learning outcomes and quantifying success across multiple disciplines, student bodies, and instructors. In examining why some online classes succeed while others fail, we must look both at their students and their production methods, rather than treat all online courses as if they are the same.

REFERENCES

1. Deborah F. Carter. 2006. Key issues in the persistence of underrepresented minority students. *New Directions for Institutional Research*, *2006*(130), 33-46.
2. Jason Caudill 2017. The Emerging Formalization of MOOC Coursework: Rise of the MicroMasters. In *EdMedia: World Conference on Educational Media and Technology*, 1-6. Association for the Advancement of Computing in Education (AACE).
3. Binglin Chen, Matthew West, and Craig Zilles. 2017. Do Performance Trends Suggest Wide-spread Collaborative Cheating on Asynchronous Exams? In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale,* 111-120. ACM.
4. Kimberly Colvin, John Champaign, Alwina Liu, Qian Zhou, Colin Fredericks, and David Pritchard. 2014. Learning in an Introductory Physics MOOC: All

Cohorts Learn Equally, Including an Online Class. *The International Review of Research in Open and Distributed Learning, 15*(4).

5. W. Timothy Coombs & Sherry J. Holladay. 2006. Unpacking the halo effect: Reputation and crisis management. *Journal of Communication Management*, *10*(2), 123-137.
6. Susan Dynarski. 2018, January 19. Online Courses Fail Those Who Need Help. *The New York Times*, BU3.
7. Joshua Goodman, Julia Melkers, & Amanda Pallais. 2016. Can Online Delivery Increase Access to Education? (No. w22754). National Bureau of Economic Research.
8. Mark Guzdial. 2018, January 5. "Learning Computer Science is Different than Learning Other STEM Disciplines" [Blog Post]. In Blog@CACM. Retrieved from http://bit.ly/2DqCyV1
9. Ashok Goel & David A. Joyner. 2016. An Experiment in Teaching Cognitive Systems Online. In Haynes, D. (Ed.) *International Journal for Scholarship of Technology-Enhanced Learning 1*(1).
10. Ashok Goel. 2018. Preliminary Evidence for the Benefits of Online Education and Blended Learning in a Large Artificial Intelligence Class. In Madden, A., Margulieux, L, Kadel, R., & Goel, A. (Eds.) *Blended Learning in Practice: A Guide for Practitioners and Researchers*. MIT Press.
11. Khe Foon Hew & Wing Sum Cheung. 2014. Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, *12*, 45-58.
12. David A. Joyner 2017. Scaling Expert Feedback: Two Case Studies. In *Proceedings of the Fourth ACM Conference on Learning @ Scale,* 71-80. ACM.
13. Joyner, D. A. (2017). Congruency, Adaptivity, Modularity, and Personalization: Four Experiments in Teaching Introduction to Computing. In *Proceedings of the Fourth Annual ACM Conference on Learning at Scale*. Cambridge, Massachusetts. ACM.
14. David A. Joyner. 2018. Intelligent Evaluation and Feedback in Support of a Credit-Bearing MOOC. In *Proceedings of the 19th International Conference on Artificial Intelligence in Education.* London, UK. ACM.
15. David A. Joyner. 2018. Building Purposeful Online Learning: Outcomes from Blending CS1301. In Madden, A., Margulieux, L, Kadel, R., & Goel, A. (Eds.) *Blended Learning in Practice: A Guide for Practitioners and Researchers*. MIT Press.
16. René F. Kizilcec & Sherif Halawa. 2015. Attrition and achievement gaps in online learning. In *Proceedings of the Second ACM Conference on Learning @ Scale,* 57-66. ACM.
17. Steve Kolowich. 2013. A university's offer of credit for a MOOC gets no takers. *The Chronicle of Higher Education*.
18. Steve Kolowich. 2013. SUNY signals major push toward MOOCs and other new educational models. *The Chronicle of Higher Education*.
19. George Lăzăroiu, Gheorghe Popescu & Elvira Nica. 2016. The Feasibility of Coursera as a Platform for Credit-Bearing Courses. In *Proceedings of eLearning and Software for Education 3*(46), 46-52.
20. Anne E. Marshall. 2017. "A Preliminary Assessment of an MIT Campus Experiment with an edX Online Course: The Pilot of 6.S064 Circuits and Electronics." MIT Teaching+Learning Lab. Retrieved from http://bit.ly/2DwpDFp
21. Fred Martin. 2013. Fight the MOOC-opalypse! and Reflections on the Aporia of Learning. *Journal of Computing Sciences in Colleges*, *28*(6), 5-6.
22. Patrick McGhee. 2012. "Why online courses can never totally replace the campus experience." *The Guardian*. Retrieved from http://bit.ly/2DyUmRi
23. Laura Pappano. 2012, November 4. The Year of the MOOC. *The New York Times* (p. ED26).
24. Miranda C. Parker, Mark Guzdial, & Shelly Engleman. (2016). Replication, Validation, and Use of a Language Independent CS1 Knowledge Assessment. In *Proceedings of the 2016 ACM Conference on International Computing Education Research*. ACM.
25. James W. Pennebaker, Samuel D. Gosling, & Jason D. Ferrell. 2013. Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *PloS one*, *8*(11), e79774.
26. Stefan A. D. Popenici & Sharon Kerr. 2017. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, *12*(1), 22.
27. Justin Reich. 2015. Rebooting MOOC research. *Science*, *347*(6217), 34-35.
28. Ry Rivard. 2013. "Citing disappointing student outcomes, San Jose State pauses work with Udacity." *Inside Higher Ed.* Retrieved from http://bit.ly/2F3QvZ
29. Stephen Ruth. 2014. Can MOOCs Help Reduce College Tuition?: MOOCs and technology to advance learning and learning research (Ubiquity symposium). *Ubiquity*, *2014*(July), 3.
30. Jason E. Stone. 2016. Awarding college credit for MOOCs: The role of the American Council on Education. *Education Policy Analysis Archives 24*.
31. Carl Straumsheim. 2015. "323 learners eligible for credit from MOOCs at Arizona State U." *Inside Higher Ed*. Retrieved from http://bit.ly/2n3zYy