Squeezing the Limeade: Policies and Workflows for Scalable Online Degrees

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

ABSTRACT

In recent years, non-credit options for learning at scale have outpaced for-credit options. To scale for-credit options, workflows and policies must be devised to preserve the characteristics of accredited higher education—such as the presumption of human evaluation and an assertion of academic integrity—despite increased scale. These efforts must follow as well with shifting from offering isolated courses (or informal collections thereof) to offering full degree programs with additional administrative elements. We see this shift as one from Massive Open Online Courses (MOOCs) to Large, Internet-Mediated Asynchronous Degrees (Limeades). In this work, we perform a qualitative research study on one such program that has scaled to 6,500 students while retaining full accreditation. We report a typology of policies and workflows employed by the individual classes to deliver this experience.

Author Keywords

Online degrees; accreditation; academic integrity

ACM Classification Keywords

• Applied computing~Computer-assisted instruction • Applied computing~Distance learning • Applied computing~E-learning

INTRODUCTION

Recent efforts in learning at scale have focused (though not exclusively) on Massive Open Online Courses (MOOCs) and similar non-accredited offerings. Where analysis is performed within a for-credit context, research typically leverages existing large datasets (e.g. [2]) or focuses on smaller studies that promise future scaling (e.g. [8, 19]).

These efforts are all highly valuable and play an important role in expanding access to high-quality education worldwide. However, education serves multiple purposes: learning is one, but endorsement of learned skills is another.

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

ACM ISBN 978-1-4503-5886-6/18/06…$15.00
https://doi.org/10.1145/3231644.3231649

Endorsement for college credit carries several expectations and regulations not generally applied to non-credit offerings, such as requirements of FERPA compliance [20], presumption of human evaluation and feedback [10], and assertions of academic integrity [3]. To scale for-credit education, these stipulations must be included.

To investigate how to apply learning at scale to accredited offerings, we investigated an online for-credit Master of Science in Computer Science program. This program possesses full accreditation as an equivalent Master's degree to the on-campus program and has scaled to 6,800 students since its inception within the past five years. Employing a qualitative methodology, we interviewed the head teaching assistants within the program for details regarding the administration of individual classes and synthesized this with our own prior knowledge of some of the administrative procedures of the program. In this paper, we present the defining characteristics of what we call Limeades (Large Internet-Mediated Asynchronous Degrees), then present the results from our investigation into the policies and workflows one program and its constituent courses employ. These results are presented as a typology of common approaches to resolving the challenges of teaching material and endorsing knowledge at scale.

KEY PROGRAM CHARACTERISTICS

There are several key differences that separate the program we are analyzing both from traditional distance learning offerings and from MOOCs. First, differentiating the program from MOOCs is its status as a fully accredited degree. This carries two significant implications: first, the courses are subject to all for-credit standards regarding human feedback, academic integrity, and reliable evaluation. Second, as a full degree rather than a collection of isolated courses, the program demands higher-level procedures for admissions, tuition, registration, and grades.

Despite this, the program possesses many of the common features of MOOCs in contrast to older distance learning methods. The program is low-cost at between $6100 and $7100 for the entire degree; a 16-week (one semester) class costs $510, comparable to most edX MicroMasters offerings. Like most MOOCs, the program is asynchronous and geographically distributed with no contemporaneous or collocated requirements. Courses follow a cohort model with in-semester deadlines, but we observed all lectures, assignments, or tests were available for four or more days.

The combination of these factors has made the program very large: in Spring 2018, 6,500 unique students enrolled accounting for 8,900 individual class enrollments. Additional teaching assistants are hired as enrollment rises: 260 teaching assistants were hired to meet the demands of these 8,900 enrollments. The average course completion rate is 85%, so there is little difference between the enrollment and completion rates.

Colloquially, we refer to programs like this as Limeades: **L**arge **I**nternet-**Me**diated **A**synchronous **De**grees. We believe that low tuition is necessary for a program like this to become large, and that geographic distribution is an effect of a program being asynchronous and Internet-mediated (although majority-online programs with in-person exams or labs exist). It may be that not all of these characteristics are necessary to scale, but all five Limeades of which we are aware possess all these characteristics. Our research here focuses on the policies and workflows that are employed by one of these programs to scale while preserving its status as an accredited degree.

RELATED WORK

Although no universal definition exists for the 'massive' part of MOOC, we differentiate 'massive' and 'large' in terms of the relationship between enrollment and human resources: a 'massive' course is one in which the human resource allocation is constant regardless of enrollment, while a 'large' course is one in which the human resource allocation scales linearly with enrollment and where there is no cap to this growth (e.g. lecture hall size).

Although we draw a fundamental difference between MOOCs and Limeades, research in the MOOC space has addressed concerns applicable to for-credit offerings. For example, asynchronous classes often rely on exams that are open for several days, raising concerns that students may share information. This may be addressed via thoughtful test design [11], but research also shows that there is little evidence that such information-exchange exists or is having an effect [2]. Similar work exists for improving the efficiency of human graders [8], detecting and preventing cheating in semi-asynchronous courses [13], and integrating MOOCs with traditional education [4, 16].

Work has been done on scaling for-credit graduate education online [10]. That work, though, emphasized the backgrounds and motivations of human teaching assistants. Here, we focus instead on the policies and workflows that are used to run large graduate-level online classes.

METHODOLOGY

The primary focus of this analysis is on the administration of individual courses at scale in the program, with the overall degree providing background into the constraints on administering those courses. As such, this investigation is based first on a series of interviews conducted with the head teaching assistants (TAs) in the program. Of the 29 classes in the program, 25 have head teaching assistants. These 25 teaching assistants were invited to participate in synchronous (via video conference) or asynchronous interviews (via email, with an initial batch of questions and follow-up questions given based on the answers). 24 responded, with 15 selecting asynchronous interviews and 9 selecting synchronous interviews. Of these 25 respondents, 9 were also current students in the online program and 4 were alumni in the program; thus, their interviews also touched on their perspectives as students. The remaining 11 were 6 on-campus PhD students, 4 on-campus Master's students, and 1 on-campus employee. In addition, 3 of the remaining 4 classes are administered by the authors, and so our personal knowledge of the details of our classes is included in this analysis. All subjective evaluation is derived from interview participants; only objective details of the authors' classes are included.

After concluding these interviews, responses were summarized and placed into a spreadsheet for analysis. Most questions are objective, while subjective reflections are reported without interpretation prior to the discussion section. We also collected and included course syllabi in our analysis. Program-level details are abstracted from knowledge the authors have as part of working on their own classes and details that arose during the interviews.

PROGRAM DEMOGRAPHICS

A complete analysis of the demographics of the program can be found in [9] and [10]. This section instead provides background that will be relevant to interpreting and evaluating the policies and workflows described below.

As noted, the program under analysis is offered at a low cost and requires no required synchronous or collocated activities. As a result, students in the program are approximately 10 years older and far more likely to be employed full-time and have families than students in the on-campus program. This introduces scheduling challenges when it comes to features like office hours or team projects, which may benefit from synchronous interaction even while leaving it optional. It also, however, leads to a more diverse set of student experiences than the on-campus program, including students employed in a variety of fields; this motivates facilitating student-student interaction as a means of enhancing the program. Interestingly, the online program draws a much greater domestic audience than the on-campus program, which is primarily international students.

PROGRAM-LEVEL POLICIES AND WORKFLOWS

Numerous policies and workflows exist to allow this program to function. For this analysis, we are focusing specifically on the in-course learning experience, especially as driven by teaching teams within individual courses. Thus, departments like admissions, advising, and student affairs—while present in the program—are outside the scope of our analysis here.

As it applies to the student learning experience, the most relevant portion of the program-level workflow is the creation of teaching teams and classroom rosters. As an equally-accredited program, the online program must

follow the same registration calendar and procedures as campus, but unique requirements alter the online process. Whereas on-campus classes can set class sizes based on room capacity, the limitation on the online program is the number of available TAs. This creates a causality dilemma where enrollment numbers dictate TA hiring, but TA hiring limits enrollment capacity. Thus, the workflow begins with soliciting applications for teaching assistants 2-3 months before the semester starts (one month before registration starts). Applicants list and prioritize class teams they may want to join. These lists are then pivoted from individual applications to lists of applicants, which are sent to the instructors, who may add, remove, or reorder names, then return the names to the advisers for hiring.

Prior to the start of registration, the advising team hires TAs for each class in line with previous enrollment patterns. Historically, enrollment has not been sufficiently consistent to accurately predict total demand, and so this initial hiring targets the lowest expected enrollment. When registration begins, students can register for up to 3 classes. Once a class fills to its initial capacity, students join the wait list. When enough join to justify hiring another TA, a TA is hired and the capacity is lifted. One TA is hired for every 50 enrolled students, in addition to one head TA per class.

In practice, this workflow is somewhat chaotic. First, many TAs apply to work for multiple classes, and often their less-preferred class may need a new hire before their more-preferred class; this makes pairing TAs with classes difficult. Second, institute rules dictate that first-semester students may not enroll until a later phase of registration, typically only days before the semester. Thus, an accurate understanding of class demand is unavailable until right before the semester starts. Third, students—wary of the non-guaranteed status of their position on the wait list—often join multiple wait lists to ensure they get *a* class, and thus perceived demand is far higher than actual demand. Nearly every class ends registration with open seats, even though the number of seats offered is set to match demand.

Although these details are largely procedural, they reflect notable challenges to scaling for-credit education. The program is largely constrained by systems developed for on-campus offerings, such as registration calendars and semester deadlines. TA hiring requires mail-delivery of paper forms to human resources. The enrollment system assumes physical classrooms will cap enrollment. Although some challenges could be solved with custom tooling and policies, it is often adherence to existing procedures that allows a new delivery medium to retain equal accreditation.

ASSESSMENTS, EVALUATION, AND FEEDBACK

Chief among the concerns in delivering a scalable online degree is supplying human evaluation. To remain comparable to the on-campus degree program, this must include both human-generated endorsements of student ability and human feedback to help improve student knowledge and correct misconceptions.

All classes in the program have assignments or projects; no class evaluates students based on tests alone. For many classes, these are programming projects, but other classes require essays, data analysis, prototypes, and other open submissions. Portions may be programmatically evaluated, but no class in the program relies solely on autograding.

Manual Feedback

At large class sizes, managing grading workflows to ensure feedback and grades are prompt, accurate, and consistent presents significant challenges. Most TAs say they aim for a one-week turnaround time on grade return.

Grading Workflows

Through our interviews, we identified four common grading workflows. We label these Randomized, Sectioned, Segmented, and Individual. Randomized and Sectioned are the most common, while Segmented and Individual are used sparingly. Some classes use different workflows for different assignments, specifically a Segmented approach for exams alongside other workflows for other assignments.

In a Randomized workflow, graders are assigned to students randomly for each assignment. Graders typically enter grades into a collaborative spreadsheet, and the head TA exports the collaborative spreadsheet into the learning management system (LMS). Interestingly, multiple TAs noted that they felt this workflow should be (but is not currently) native to the LMS (first Sakai, then Canvas).

In a Sectioned workflow, each TA is grades the same section of students all semester. This better distributes accountability by giving each student on point of contact; it lets graders develop a familiarity with individual students; and it minimizes weekly grading setup. TAs noted drawbacks, however: it limits perspectives that students receive; it hinders intergrader reliability; and it may hurt students assigned to a subpar or harsher grader.

In a Segmented workflow, individual assignments are broken into parts, and each grader grades the same part for the entire class. TAs reflected this aids with intergrader reliability; however, it is procedurally difficult to do and only fits decomposable assignments.

In an Individual workflow, one grader grades every student. The smallest classes in the program use this workflow by default as they only have a couple graders. Two larger classes reported using this workflow, but in their cases each grader grades a different assignment. This also aids intergrader reliability and preserves multiple perspectives, but it creates radically imbalanced workloads over time.

Workflow selection is based in part on the assignment structure and in part on the instructor's priorities. Sectioned workflows reflect priorities of building rapport and community, while Randomized workflows emphasize consistency and reliability.

It is worth noting that all four of these workflows have been present in on-campus classes in the past. The noteworthy

takeaways are that such general workflows can be used as-is at scale, but that large class sizes put added pressure on systematizing the process and minimizing edge cases.

Regrades

Most classes accept regrade requests, but policies regarding regrades are varied. Some classes only accept regrades unofficially; TAs reflected that providing a prescribed route for regrades leads to more requests. Other classes advertise a form to fill out to more heavily systematize requests. Upon receipt, most classes assign the request to the original grader, but some deliberately assign it to a different grader.

Intergrader Reliability

Many TAs are concerned with whether multiple graders grade consistently. TAs reflected this takes on added emphasis in large classes because (a) large teaching teams raise the likelihood of some inconsistency, (b) distributed teaching teams may not automatically converse as much, and (c) students are more likely to post public grading questions, highlighting inconsistency.

We identified three general categories of approaches to asserting intergrader reliability: Systematic, Prospective, and Retrospective.

Systematic checks are those built into the structure of the grading workflow. In the typology of grading workflows above, Segmented and Individual inherently give assert reliability. Systematic approaches can also be integrated into other workflows: some classes that use Randomized workflows assign certain students to multiple graders, then use the results from those "shared students" to identify instances of inconsistency. These inconsistencies are then resolved mathematically.

Prospective approaches attempt to resolve intergrader reliability before it has occurred. For many classes, this comes from using formal rubrics, although TAs of classes with more open-ended assignments and projects note that writing such rubrics is difficult. Other classes take a more ad hoc approach and encourage graders to converse frequently before and during grading about common edge cases. Some classes follow the "shared student" workflow above, but have all graders grade the shared students first so inconsistencies can be identified before grading begins.

Retrospective approaches identify inconsistency after grading has finished through a separate process. Many classes do so mathematically, taking a general average of the graders and applying some correction; TAs for these classes argue that with random assignment of 40-50 students to each grader, it is unlikely that a grader gets a truly subpar group overall, and thus averages should be similar. One class has each grader review another grader's work for inconsistency with their own approach.

Automated Feedback

Roughly half the classes in the program use some form of automated grading and feedback. These are commonly used on programming assignments or projects where either success can be measured by passing test cases or where performance can vary and be tested systematically. We observed two primary approaches to autograding, which we label Synchronous and Asynchronous.

Asynchronous

Asynchronous autograding, which we also call 'batch' autograding, occurs after the submission deadline. All student submissions are downloaded and local testing scripts examine them one-by-one, generating grades and feedback which are then batch-exported back to the students. In most cases, batch autograding is one component of the project or assignment grade; human graders are then equipped with the results from this batch autograding to use during the manual portion of grading, which often focuses on finding partial credit opportunities.

Synchronous

Synchronous autograding allows students to submit their code to an autograding server and receive live feedback. This approach is most commonly used in artificial intelligence and machine learning classes where students develop code using a training set that they can see, and are graded on its performance against a test set that they cannot see. This also lets students verify that their code meets requirements for function names and output formatting prior to grading. TAs reflected that this has significant pedagogical value in supporting revision and rapid feedback cycles, but that the cost of implementing such a system is high: they note it is only because of the number of students it supports that classes can justify the cost.

Group Projects

Many classes in the program require group projects, in contrast to the common view that groupwork is incompatible with online education. Group communication is handled by the groups themselves, and so one of the major questions for how to administer such projects is outside of the scope of this paper. For the teaching teams, however, team assignment is a challenge.

The three mechanisms we identified that classes use for team assignment are not remarkably different from those on campus: Self-Selected, Randomized, and Systematized. A Self-Selected workflow allows students to form their own groups; those that do not are typically assigned Randomly or Systematically, although TAs remark that these students are often disadvantaged as self-formed groups are usually more engaged in the class. Random assignment to assign students with no information affecting the process.

Most classes use a Systematized approach, where they solicit information from students and attempt to use that information to form groups intelligently. Among the information solicited are technical background, subject interest, time zone, and desired meeting time. Notably, however, TAs reflected that as soon as they try to optimize based on 2 or 3 of those components, they find a significant number of students with no pair. Students also regularly comment with different concerns about their team,

introducing an additional layer of complexity where each characteristic not only has a nominal value, but also a priority level for each individual student. Thus, team formation remains a promising area for more research.

STUDENT-TEACHER INTERACTION

Human evaluation is one element of interaction that must be preserved in a for-credit program, but for the online experience to be comparable to the in-person experience (as it must be to possess equivalent accreditation), other forms of student-instructor interaction must be offered as well. In our interviews, two major forms of this arose: interaction on forums and interaction through office hours. In addition, several classes have explicit participation policies aimed at fostering interaction among students as well.

Forum Interaction

Forum participation in online classes has long been a topic of research [7], with even more work emerging due to the prominent nature of forums in MOOCs [e.g. 1, 14, 15]. Most of this research focuses on valuable social learning experiences among students. For this analysis, however, we are concerned with forums for official communication between students and teaching teams.

All TAs reflected that the course forum (typically Piazza) is a major hub of activity for the class. Several TAs noted something to the effect of the forum being the virtual "classroom" rather than the video lectures. Under this model, the lecture videos fill the functional role of a textbook, but the forum supports the actual interaction.

Question-Answering

Within the forum, students have an expectation of official answers to questions. TAs note this expectation comes in part from the role of tuition in assuring students they are "real" students; in part from the promise that the online program is equivalent to the on-campus program; and in part from the high stakes attached to course grades. TAs suggest that the persistently available nature of the online forum also raises students' expectations on response rate; they note that if a question goes 24 hours without a response, students begin to comment. At the same time, courses can draw a high question volume. A typical class may draw 10 threads and 100 replies per day, while carrying this expectation of rapid and official responses.

Thus, course teams develop workflows of ensuring students receive timely answers. Based on descriptions from the TAs, we group these workflows into four major categories: Designated Forum TA, TA-Per-Day, TA-Per-Topic, and Free-for-All. TAs note that the goal of these workflows is generally two-fold: one, to get students rapid, accurate responses, and two, to create accountability so head TAs may identify who was responsible for an unanswered post.

Most classes take the Designated Forum TA approach. In this approach, a specific TA—often the head TA, but sometimes a designated other TA—is solely responsible for answering forum threads. This approach is most commonly seen in classes that either have a particularly conscientious TA who is comfortable frequently checking the forum, or with small classes with manageable question volumes.

Head TAs for other classes report that forum administration is too large a task for a single person, so they divide the responsibility. Most use a TA-Per-Day workflow, where each TA is assigned a day, and their responsibility is to address all unanswered posts by the end of their day. A handful of classes take a similar TA-Per-Topic approach, dividing by topic or folder; each TA answers a certain type of question, often those for assignments they wrote.

Lastly, a few classes take a 'Free-for-All' approach. In the interviews, these were those head TAs who did not have an official process for ensuring questions are answered. In some cases, their reply to the question was effectively, 'We're all attentive and responsive', while others replied, 'The instructor handles them and sometimes we help.'

Notably, multiple TAs (although a minority) noted they intentionally delay their answers to give students the opportunity to answer each other's questions. They argue this fosters community as well as makes the TAs' job easier. Other TAs argued that rapid responses make students more likely to visit the forum frequently, which leads to greater student-student interaction as a byproduct.

Announcements

TAs noted course announcements play a significant role in their courses; while the content of the announcements is important on its own, announcements also establish a cadence within the course. Such a routine is usually supplied by the lecture schedule, but is absent online.

To address this, several classes use weekly announcements. These announcements usually have no information that cannot be found elsewhere: they restate what lectures to watch for the week, what assignments are upcoming, when office hours can be expected, and so on. Despite the lack of any unique information, students reflect that these are one of the most-appreciated things that classes may choose to do. It functions in part to establish a community cadence to replicate that effect of scheduled lectures, but it also changes the information retrieval dynamic from 'pull' to 'push'. Rather than students relying on the fact that they have successfully 'pulled' all relevant information from the course web site at the right time, they are instead reassured that important information will be 'pushed' to them.

To further create a cadence or routine, some classes extend this to daily threads, like Monday Reflections and Friday Humor, to seed student participation on the forum.

Office Hours

All 27 classes covered by this study offer some form of office hours. We identified three different types of office hours, which we labeled Traditional, Supplemental, and Textual. Traditional and Supplemental dominate in approximately equal portions, while four classes use Textual office hours. Most classes use video, but there is variation within that; one class screencasts office hours

from a tablet so they may write, while others broadcast either an individual or a roundtable with multiple people.

Nearly all classes offer office hours with the instructor and the TAs. Frequency of office hours vary wildly, from ~5 times per semester to ~8 hours per day all semester. This is largely determined by class enrollment; large classes have large TA teams capable of offering more constant office hour sessions. The amount of student attendance varies as well; some TAs noted they rarely get more than one or two attendees to a session, while others noted a typical session would draw 40 or 50. We hypothesize this is a function of the style of office hours, the nature of the course material (e.g. technical or theoretical), and the responsiveness of the teaching team via forums.

Traditional

We defined 'traditional' office hours as sessions that mirror the common construct of a student talking 1:1 with a professor or TA privately or semi-privately in an office or common area. Most classes in the program offer this form of office hours. Schedules are provided to students via mutable interfaces like Google Calendar or immutable interfaces like the syllabus. TAs reported using Webex, BlueJeans, or Google Hangouts for these office hours.

An interesting phenomenon many TAs noted was students "sitting in" on office hours. In these instances, students would sign into office hours just to listen, not to ask their own questions. Several TAs reflected negatively on this phenomenon, believing it disturbed the semi-private nature that office hours are supposed to carry. One TA noted he uses Webex's "waiting room" function to let students into office hours one-by-one to compensate. Multiple TAs similarly remarked that students complain that their office hours are not recorded and shared with the class; they noted that this similarly breaks the semi-private nature.

Supplemental

Upon reflection, we believe that these behaviors and complaints from students are because other classes in the program offer what we describe as 'supplemental' office hours. In supplemental office hours, instructors and TAs solicit questions from students in advance, then answer them live through YouTube Live or BlueJeans. Recordings are then shared with the class. In some cases, questions are solicited live from a chat interface, but the conversation is not live between the teaching team and the students.

We describe these office hours as 'supplemental' because we perceive these alternative office hours as providing students a venue to request supplemental lecture material rather than have a live conversation. We hypothesize that the negative behaviors observed in "traditional" office hours are due to expectations set in classes that offer "supplemental" office hours.

A few classes offer both Traditional and Supplemental office hours. The instructor typically offered supplemental office hours, while the TAs offered traditional office hours.

Textual

The Textual approach differs first in medium rather than style; however, TAs reflected that this altered medium leads to a change in style. These classes use the workspace chat tool Slack to offer chat office hours. For some classes, the teaching team schedules times when they pledge to respond immediately on Slack; others give a more general note that the teaching team is active, and so to contact them any time.

We categorize this approach separately because it captures elements of both Traditional and Supplemental office hours. Like Traditional, students may spawn private chats; but like Supplemental, chat logs in public rooms are persistent for students to read through, like a recording. TAs observed that shifting from traditional office hours to textual office hours has given rise to more casual interaction as well as more student-student interaction in public chat rooms during office hours. We hypothesize this is due to the lower technical and social barriers to participation.

Class Participation

To replicate the type of class discussion experience present in some on-campus classes, several classes in the program have formal participation requirements. At a minimum level, in some classes this is a requirement to post a certain number of times to the forum (either anywhere or in designated topics) or to read a certain proportion of the class's posts. TAs who had taken classes with these policies reflected a general distaste for them, feeling it made the discussion inauthentic.

Two classes have far more complex participation policies. These classes, run by the same instructor, require students to earn participation points, but they provide multiple avenues to maximizing credit, such as peer review, forum participation, beta testing classmates' projects, or completing classmates' surveys. The purpose stated in the syllabus is to have every student spend 1 hour per week contributing to making the class better; with 17 weeks and 250 students, 4,000 hours are spent enhancing the class.

POLICIES

Every course has several policy decisions to make regarding plagiarism, collaboration, anonymity, and privacy. As part of our interviews, we gathered syllabi from TAs and examined policies within these categories, as well as asked TAs to report enforcement processes.

Plagiarism & Collaboration

Most classes have the same general plagiarism and collaboration policy: students are permitted to talk at the "whiteboard level", as multiple TAs said, but should avoid copying code or text. Some classes have explicit policies on copying code, noting that students may copy smaller utility functions or test cases; these classes explicitly note that students must cited such copied code. TAs noted that the presence of citation is the difference between a poor grade and a student integrity violation: if a student copies too much but cites what they copy, their grade may be penalized, but they will not be accused of plagiarism.

Plagiarism Detection

The program provides plagiarism-detection tools and encourages systematic usage within courses. These largely focus on detecting code plagiarism, which TAs noted is a greater need due to perceptions about the acceptability of copying code and the difficulty in diagnosing individual cases. One TA noted that the scale of the online program allows resources to be invested into autograding software, but this demands project re-use between semesters, raising the likelihood of plagiarism. Other TAs have developed additional tools and policies to deter or detect plagiarism; one notes such efforts reduced plagiarism from 14% to 2%.

The tool compares all student code submissions against all current and previous submissions to the same project deliverable. In this way, it checks for both disallowed collaboration (shared code between students in the same semester) and copying (code borrowed from a previous student). Cases of plagiarism follow the institute's rules for student integrity, which are difficult to follow at scale as they demand meetings with each accused student.

Anonymity and Privacy

All classes also allow students to post privately to instructors; this is typically the recommended venue for grade disputes or other personal administrative questions. Instructors may also allow students to post with complete anonymity (to instructors and classmates alike) on Piazza. Most classes do not do so: TAs noted that allowing students to post with complete anonymity often leads to students writing unprofessional or rude complaints that can quickly spark a revolt in the class. Even without anonymity, defusing emotionally-charged forum threads has become a valuable skill for online instructors to learn. We recommend future research on tactics for such interactions.

Piazza always lets students posting with anonymity to classmates, however. TAs remarked that students rarely post anonymously to classmates, but when they do, it appeared to be due to fear of judgment. In some cases, they believed students feared being judged based on their question, but in others, they suspected students feared judgment based on the gender or nationality apparent in their name. One TA noted certain students always post anonymously, likely due to personal beliefs about privacy.

SWEETENING

At the end of our interviews, we asked the participants to let us know anything their classes do that they would consider unique, either within the online program or within higher education more generally. Going back to our "Limeade" acronym, we describe these things as "sweetening": features that improve the program by leveraging its scale.

In-Class Competitions

Four classes reported that they have some type of in-class competition. These classes are all artificial intelligence or machine learning courses, for which there exist significant opportunities for optimizing the performance of an agent or algorithm. Most classes rank student submissions based on performance according to some metric like performance against a test set. One class, where students design game-playing agents game, holds an automated tournament amongst the student submissions to find a "champion".

The incentives attached to these competitions vary by class: some give extra credit, while others give public recognition. TAs noted that these competitions are more interesting in large classes: "winning" in a class of 500 is a greater honor, they note, than "winning" in a class of 25.

Mentorship

Two classes are built around a pedagogical model of mentorship. In these two classes, students—individually or in teams—are partnered with mentors who evaluate and give feedback on all the student's work, developing a familiarity with them and their project history that TAs note enhances the feedback mentors give.

Notably, the mentors for both classes possess subject matter expertise. One class covers educational technology and counts among its mentors data scientists for EdTech companies, executives from textbook publishers, university professors; and AP Computer Science teachers. Here, the model relies on the program's scale to deliver enough students with have professional backgrounds in the area. The other class, on healthcare technology, typically partners students with medical doctors for developing patient- and practitioner-centered interfaces. Here, the model thrives because the unique demographics of the program mean the student groups are qualified to authentically contribute to the field, leading to benefits for students and mentors alike.

Community-Building

Several TAs noted an early concern was developing student communities within their courses. TAs also reflected that this concern was ultimately unfounded; communities formed on their own through the course forums and other social networking services. However, several initiatives were originally devised to improve student communities.

Peer Review

Roughly one-third of the classes in the program use some form of peer review. In contrast to MOOCs, however, no classes use peer review for generating grades. Instead, TAs report pedagogical benefits of peer review. First, they remark students benefit from seeing alternative approaches. Second, they note that students benefit from putting themselves in the role of a grader in critiquing others' approaches. Third, they suggest being aware of classmates' works increases the feeling of being part of a broader class community, like sitting together in lecture on campus. Fourth, they believe students benefit from receiving more feedback than they would from a grader alone; however, some TAs remarked that as students, they rarely found the peer reviews they received to be particularly insightful. They felt the first three benefits were more significant.

Assignment Libraries

In one class, students propose their own projects to complete within the subject matter of the course. As such,

every student's assignments and project are different, but they all follow a standard trajectory of initial research, proposals, milestones, and final presentations. This course developed the idea of an Assignment Library, where students who opt-in have all their submissions shared with the rest of the class. In this way, students who come across interesting work during peer review can watch that work develop over time. Students also peruse the library for similar work to use in team formation, and to follow along with similar projects to benefit from their classmates' research. This class shares final papers and presentations from previous semesters with subsequent semesters so students may build on prior students' work and develop a broader community of practice. Although this approach is compatible with small courses, the size of the course raises the likelihood that students will find relevant projects.

Merging Online and On-Campus

A handful of classes have experimented with merging online and on-campus forums. TAs of classes that have tried this report mixed but generally positive results. The most positive reflections on this idea come from classes where the online and on-campus demographics complement each other well; in the aforementioned healthcare technology class, for example, the online students often have more professional software development experience, while the on-campus students have more courseware in interaction design. Negative reflections largely came when a shared forum highlights features that were accessible to only one audience, such as on-campus students discussing in-person lectures that were unavailable to online students.

Course Hand-Offs and Co-Instructors

Finally, one interesting dynamic mentioned by multiple TAs was the notion of a course "hand-off". On campus, it is not uncommon for a class to be taught by a different instructor in different semesters, or for courses to be passed from instructor to instructor over many years.

Course Hand-Offs

When online instructors retire, change schools, or wish to stop teaching, their recorded lectures can persist. Delivery is handed off to another instructor who answers forum questions and runs course delivery.

TAs had mixed reactions to this. Some noted a cognitive dissonance occurs when taking a course without the original recording instructor. In some cases, the new instructor seemed ill-suited to the original instructor's teaching style. In others, however, TAs noted that they benefited from receiving two different views in the same class. Other TAs noted some instructors excel at forum administration while others excel at lecturing, and these hand-offs allow the varying roles associated with class administration to be filled with ideal candidates rather than forcing a single instructor to fill all roles at once.

Teams of Co-Instructors

This latter view has led to a couple classes experimenting with co-instructors. In these classes, one co-instructor (often the original instructor who recorded the class) is the content expert, while another handles class administration. The content expert focuses on answering interesting content questions, while the administration instructor takes care of TA teams, grade workflows, and the other components described in this analysis. This allows the original instructor to remain involved with less pressure to run every element of the class. It also increases the pool of possible co-instructors as the class administration instructor needs not have as much subject matter expertise. One TA noted that such an arrangement is uniquely possible online: the environment allows the instructors to work in parallel more easily, while the increased class size justifies having multiple instructors assigned to a single course.

CONCLUSION

In this work, we have analyzed the policies and workflows that have allowed an accredited Master of Science program to scale to almost 9,000 seats online. The most significant conclusion of this work, however, is likely what *can* be done in this space. For the program to be equally accredited, the online experience must be comparable, and so several features that are often excluded from online programs for the sake of scale were deliberately included. Group work, class participation, plagiarism detection, and human evaluation are among those elements that are necessary for the online program to remain comparable, and so the program found ways to deliver these online.

Beyond recreating these features, in many ways the scale of the program improved them. TAs noted that a class of 500 students gives a greater likelihood that students will find teammates with related interests. Huge submission numbers for assignments means intergrader reliability evaluations take on greater rigor as they act on large sample sizes. With forum- and chat-based student communities, students can typically find a classmate online at all hours; large TA rosters to meet large class sizes also mean that office hours can be made available nearly all the time.

In our interviews, TAs observed few sacrifices in classes within the program for the sake of scale. One could argue that the sacrifices are present outside the individual classes, and that only those classes that are compatible with an online environment are converted; however, classes in the program range from lectured-based classes to discussion- and project-based classes, leaving few classes that do not present potential for online delivery.

TAs only remarked on two potential advantages that they perceive on-campus students may have. First, multiple TAs noted that the online program is coursework-only, while on-campus students may complete research projects for course credit. This is inaccurate, however: online students (including some of the TAs interviewed) have completed independent study research classes, Master's projects, and Master's theses. These options are not heavily advertised due to difficulties in scaling, leading to the perception among TAs that they are unavailable online, but they are

available to ambitious online students. Secondly, TAs noted that on-campus students have the advantage that they have access to the online lecture material *and* traditional lectures, effectively doubling the amount of material available. This advantage is interesting because it only exists due to the online program's existence.

7 Principles for Leveraging Scale

As noted, the major weakness in this evaluation is the lack of evaluation. In some ways, the evaluation actually occurred prior to this analysis: the reason this case study was worth pursuing is *because* the evaluation of the program was positive, in that it achieved scale [10] while preserving learning outcomes [4] and is projected to significantly impact the number of MSCS graduates generated in the United States each year [6].

Much of this work has focused on how the problem of scale was solved, but the more impactful observations are those about how scale can be used. Scale presents challenges, but it also presents opportunities. Therefore, we conclude with a set of principles derived from this work on how to not only solve scale, but to use scale to our advantage:

1. **Shift the investment forward.** Understand that developing an online course or program pools the resources the course will generate over several semesters into one initial investment. Components like complex custom autograders, original high-quality videos, or expensive hardware will be reused semester over semester, cutting grading and teaching costs.
2. **Queue up reusable material.** Reusing lecture videos is a natural part of teaching online, but components like assignment feedback, forum answers, and announcements can be constructed for reuse, too. Tools like GradeScope [18] and Jill Watson [5] may be used to generate high-quality reusable feedback given the knowledge that each error is likely to occur dozens of times over multiple semesters. Similarly, plagiarism-detection tools like MOSS-TAPS [17] benefit from repeated use over multiple semesters.
3. **Utilize the student body.** The lack of constraints geographical, temporal, or capacity constraints in online programs leads to more diverse student populations than traditional programs. Build in dedicated ways, like team projects, peer review, or forum participation, in which these students can formally and deliberately interact with one another and, thus, improve the courses for their classmates.
4. **Embrace the 24/7 classroom.** These classes have mechanisms to ensure student questions are answered at all times on all days because the online classroom never closes. Emphasizing the universal availability of the classroom invokes a belief in students that their community is always available, connecting them to a broader social learning experience.
5. **Emphasize iteration from the beginning.** Online courses minimize repeated work, meaning that work investment in one semester will be reused for multiple semesters in the future, leading to constant course improvement. Courses thus ought to be built specifically to allow room for iteration. This can be done by keeping the connection between course material and course assessments loose (so assessments may be revised more easily) and keeping video lessons relatively self-contained and independent for simpler addition, removal, or replacement of material.
6. **Zone off separate instructor roles.** With hundreds of students in a class, it is reasonable to hire multiple individuals to serve portions of the instructor role. These instructors can thus each perform the role at which they excel. There may be instructors without sufficient content knowledge to answer course questions, but who are experts in online course administration; at these scales, it is plausible to assign separate instruction, content, and delivery experts.
7. **Evaluate workflow alternatives.** We have uncovered workflows or paradigms for grading (Randomized, Sectioned, Segmented, Individual), intergrader reliability (Systematic, Prospective, Retrospective), forum administration (Designated TA, TA-Per-Day, TA-Per-Topic, Free-for-All), team project assignment (Self-Selected, Randomized, Systematized) and office hours (Traditional, Supplemental, Textual). All have different strengths reflect different priorities. In our discussions, however, many TAs had not considered alternative approaches. In designing course delivery, we ought to explicitly evaluate these alternatives.

The true opportunity for learning at scale comes from principles like these for using scale to improve the learning experience, not merely scaling existing learning experiences with minimal quality loss.

Limitations & Future Work

The major limitation of this study is that it is descriptive, not experimental or evaluative. It does not assert the necessity of these policies. Some may be counterproductive or undesirable; for example, research has shown that competition unfairly motivates men [12], and so leveraging competitions may exacerbate existing gender divisions.

Secondly, this study is tightly embedded within one program context. It may be that many of the structures analyzed here are unnecessary or absent in other Limeades, or that alternative policies and bureaucratic climates may make many of these workflows non-transferable.

We recommend three major directions for future work to build on the work presented here and address these limitations. First, this work and prior work has focused on evaluating the human component of the teaching side, specifically teaching assistants [10]. We elected to focus first on TAs to develop a broad view of the program; subsequent studies should focus on how the dynamics uncovered here specifically impact students.

Second, significant work has focused on the online learning experience, but instructors and TAs have noted that

developing an online course significantly impacted their residential course delivery. Many have noted that having high-quality videos alters what they do during synchronous lecture time, or that producing an online course forced them to rethink, restructure, and refresh the course content, benefiting on-campus students as well. Further research should attempt to summarize the effect online course development has on equivalent on-campus courses.

Third, there are several assumptions often stated about online learning, including that online students do not interact with instructors and peers as frequently. This has been raised as a reason why online education may not succeed at lower levels. This research indicates mechanisms are in place to facilitate such interaction, but it does not evaluate whether the quantity and quality of that interaction are comparable between online and on-campus courses. To be truly rigorous, research must compare the full range of interactions, including live question-and-answer during lecture, in-person office hours, and social "hallway chatter" immediately after lectures. We hypothesize that such an analysis would reveal a comparable amount of interaction online compared to in-person.

REFERENCES

1. Christopher Brinton et al. 2014. Learning about social learning in MOOCs: From statistical analysis to generative model. *IEEE Transactions on Learning Technologies*, *7*(4), 346-359.
2. 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* (pp. 111-120). ACM.
3. Michele Cole and Louis Swartz. 2013. Understanding academic integrity in the online learning environment: A survey of graduate and undergraduate business students. *ASBBS Proceedings*, *20*(1), 738.
4. Ashok Goel and David Joyner. 2016. An Experiment in Teaching Cognitive Systems Online. In Haynes, D. (Ed.) *International Journal for Scholarship of Technology-Enhanced Learning 1*(1).
5. Ashok Goel and Lalith Polepeddi. 2016. *Jill Watson: A Virtual Teaching Assistant for Online Education*. Georgia Institute of Technology.
6. Joshua Goodman, Julia Melkers, & Amanda Pallais. 2016. Can Online Delivery Increase Access to Education? (No. w22754). National Bureau of Economic Research.
7. Mark Guzdial and Jennifer Turns. 2000. Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences*, *9*(4), 437-469.
8. Andrew Head et al. 2017. Writing Reusable Code Feedback at Scale with Mixed-Initiative Program Synthesis. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale* 89-98. ACM.
9. David Joyner et al. 2016. Graders as Meta-Reviewers: Simultaneously Scaling and Improving Expert Evaluation for Large Online Classrooms. In *Proceedings of the Third Annual ACM Conference on Learning @ Scale*, 399-408. Edinburgh, Scotland.
10. David Joyner. 2017. Scaling Expert Feedback: Two Case Studies. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, 71-80. ACM.
11. David Joyner. 2017. A holistic approach to academic integrity in a CS1 MOOC. In *Proceedings of Learning with MOOCs IV.* Austin, TX.
12. Muriel Niederle and Lise Vesterlund. 2007. Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067-1101.
13. Curtis Northcutt, Andrew Ho, and Isaac Chuang. 2016. Detecting and preventing "multiple-account" cheating in massive open online courses. *Computers & Education*, *100*, 71-80.
14. Daniel Onah, Jane Sinclair, and Russell Boyatt. 2014. Exploring the use of MOOC discussion forums. In *Proceedings of London International Conference on Education* (pp. 1-4). LICE.
15. Justin Reich, Brandon Stewart, Kimia Mavon, and Dustin Tingley. 2016. The civic mission of MOOCs: Measuring engagement across political differences in forums. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 1-10). ACM.
16. Cathy Sandeen. 2013. Integrating MOOCs into traditional higher education: The emerging "MOOC 3.0" era. *Change: The Magazine of Higher Learning*, *45*(6), 34-39.
17. Dana Sheahen and David Joyner. 2016. TAPS: A MOSS Extension for Detecting Software Plagiarism at Scale. In *Proceedings of the Third Annual ACM Conference on Learning at Scale*, 285-288. Edinburgh, Scotland. ACM.
18. Arjun Singh, Sergey Karayev, Kevin Gutowski, and Pieter Abbeel. 2017. Gradescope: a Fast, Flexible, and Fair System for Scalable Assessment of Handwritten Work. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, 81-88. ACM.
19. Eliane Wiese, Michael Yen, Antares Chen, Lucas Santos, and Armando Fox. 2017. Teaching Students to Recognize and Implement Good Coding Style. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, 41-50. ACM.
20. Elise Young. 2014. Educational privacy in the online classroom: FERPA, MOOCs, and the big data conundrum. *Harv. JL & Tech.*, *28*, 549.