Scaling Expert Feedback: Two Case Studies

David A. Joyner
Georgia Institute of Technology
Atlanta, Georgia
david.joyner@gatech.edu

ABSTRACT
Traditionally, education relies on a linear relationship between enrollment and staff; rising enrollment dictates increases to staff with some expertise (such as teaching assistants, TAs) for evaluation. This relationship is expensive, so learning at scale has largely deemphasized expert evaluation and feedback. Two organizations, though, have used different models to scale up class size online while retaining this expert evaluation and feedback. In this paper, we analyze the methods these two organizations have used to increase enrollment while preserving scalability and feedback. We observe an academic program has scaled feedback with traditional TAs by relying on unique characteristics of its student body, while a commercial program has done so with a novel, network-based model. These successes show the potential of learning from experts at scale.

Author Keywords
Expert feedback; online higher education; microcredentials.

ACM Classification Keywords
Social and professional topics~Computer science education

INTRODUCTION
The rise of online education has been a boon for learning at scale, especially in higher education, due to the online environment's ability to resolve classical threats to scalable education, such as classroom size, geographic mobility, and synchronous scheduling. However, online delivery still may incur significant monetary costs to students because the most expensive piece of higher education is also one of its most pedagogically important: the presence of dedicated, expert-level feedback, such as that provided by by graduate teaching assistants (TAs) in many programs. Research shows students perceive these TAs as less authoritative than professors, but also as more engaging and interactive [14].

In online programs, this threat to scale has been handled in two main ways. One, many programs deliver traditional curricula online for the same cost as residential programs. The similar cost lets classical models of grading by instructors and TAs persist. These costs, however, dampen the programs' scalability due to the massive and unevenly distributed barrier to entry presented by high tuition.

So, many efforts in learning at scale emphasize *removing* individual expert-level feedback, as seen in most MOOCs. While few people would argue this is a better learning experience, some argue the trade-off in affordability justifies the shift in approach to feedback.

Recently, efforts to reintroduce expert-level feedback into affordable programs have begun. MOOC host Coursera has begun leveraging volunteer mentors to provide expertise in forum interactions [26]. While paid graduate TAs do not possess the expertise of professors, their feedback sustains existing graduate programs; scaling it could preserve expert feedback while increasing accessibility and enrollment.

This paper provides case studies on two ongoing efforts to use scalable expert-level feedback in affordable online programs. One effort is in a major university's all-online Master's degree, which enrolls over 4,000 students in its online Master's program, while keeping the *total* cost of the degree around $7,000 [4]. Importantly, the program is identically accredited as the residential program, creating demand for the same feedback and endorsement residential students receive. The university has scaled by preserving the traditional model for expert feedback, but adapting it to rely on the unique motivations of online students.

The second effort is a company providing online programs in emerging fields like machine learning, enrolling over 10,000 students, while maintaining a lower (~$200/month) program cost than accredited institutions, boot camps, or certification programs. While the company does not have accreditation requirements, it has a similar pressure to provide expert-level feedback and endorsement due to a guarantee of employment for graduates. The company has scaled using a dramatically different model for expert evaluation than traditional higher education, leveraging a distributed network of freelance professionals.

These organizations are very different. The university is a century-old accredited academic and research institution; the company is a young Silicon Valley startup. However, they experience similar pressures to provide expert-level feedback and a reliable endorsement of the skills of their graduates. So, both have developed ways to scale their programs while preserving student ↔ expert interaction.

L@S 2017, April 20 - 21, 2017, Cambridge, MA, USA

ACM 978-1-4503-4450-0/17/04…$15.00
DOI: http://dx.doi.org/10.1145/3051457.3051459

This paper first outlines reasons that methods for scaling used by other programs are insufficient given these pressures. It then gives case studies on two organizations: for each, it covers the setup for scaling feedback, and for the university, it examines the motivations of those experts. It concludes that scaling expert feedback while preserving affordability is possible.

MODELS FOR SCALABLE FEEDBACK

In examining the landscape around MOOCs, there are three common mechanisms for scaling feedback without additional hiring: automated, peer, and implicit feedback. These three require no additional investment of time from the teaching team as enrollment rises. All three also present valuable learning opportunities; there are challenges, however, in using them alone in programs with significant pressure for expert-level feedback and endorsement.

Automated Feedback

One approach to scalable feedback is automatic evaluation. Traditionally, this calls to mind negative connotations surrounding multiple-choice tests, but automated evaluation can evaluate complex problems and provide individualized feedback. Grading by simulation is one example, that has been used in fields like cyber-physical systems [11] and electrical engineering [25]. Similar initiatives, called virtual labs in the sciences, even predate MOOCs, e.g. [33]. The prevalence of online computer science education has led to initiatives for automatic code evaluation, e.g. [6, 27], although some initiatives predate online education [10].

Aside from scalability, automated feedback presents pedagogical advantages. It facilitates immediate feedback, allowing students to engage the valuable rapid revision cycles [3]. The entire field of intelligent tutoring systems [22] is built on immediate, individualized, automatic feedback [28], and research has found it can be nearly as effective as human tutoring [29], while presenting significant advantages in availability and integration.

However, automated evaluation presents three major challenges. First, construction of automated evaluators is typically expensive, a problem so apparent that dedicated efforts are underway to reduce production costs [1, 21]. Even when automated assessors can be built, equipping them with deep, expert-level feedback is an additional challenge. Second, they largely address closed answer spaces; the space may be large, but they struggle with entirely open-ended, student-driven, project-based learning, which is desirable and presents a demand for feedback [15]. Third, deserved or not, overreliance on automated evaluation may challenge the perception of the endorsement's authenticity and integrity. While automated evaluation has an important role, it cannot presently be solely responsible for feedback and endorsement.

Peer Feedback

Likely the most common approach taken to scaling evaluation online is peer feedback. Peer feedback entails having students evaluate classmates' work, then deriving a grade from those evaluations. Like automated evaluation, it is perfectly scalable: adding students also adds reviewers. Solutions for peer review are more easily constructed and domain-general, helping resolve the cost of construction.

Peer feedback has great pedagogical value. Research shows its positive effect on learning [2, 18, 31] due to the implicit feedback and learning-by-teaching paradigm; that the act of *giving* feedback itself has a positive effect on learning [20]; and that it can be valid under specific circumstances [5]. Efforts exist to improve its validity through machine learning [e.g. 17, 23] or expert meta-reviewing [e.g. 12].

Peer review is valuable; however, for supporting accredited or guaranteed programs, it presents challenges. First, as noted, it is valid only in specific circumstances, and research shows peer graders are often unreliable [32] and unmotivated [19]. Second, it also introduces perception issues: the notion that a student may graduate without any expert evaluation is problematic. While these two issues are likely resolvable with evidence of validity, a third challenge is more problematic: individual expert-level feedback provides expertise and targeted insights that peer review does not always capture. Experts' knowledge of the material translates to better feedback than peers alone. Thus, while it is pedagogically valuable to include peer review, it likely cannot solely support such high-stakes programs.

Implicit Feedback

Implicit feedback is providing resources from which learners can self-evaluate [9]. A common type of implicit feedback is supplying learners with exemplary assignments: by comparing their own work to the best work in the class, they can derive feedback on their performance. Worked examples are similar in formal domains [24, 30]. Peer review facilitates implicit feedback, too: it provides learners an opportunity to compare their work against other approaches, developing their understanding of its strengths and weaknesses.

Like peer and automated feedback, implicit feedback is perfectly scalable. It additionally may provide a richer set of feedback than the more closed set provided by automated evaluation by giving well-equipped students the data to generate detailed personal feedback. However, it presents drawbacks: it does not generate endorsement, it necessitates preexisting metacognitive skills, and it may raise more questions than it answers in the mind of the student.

Expert Feedback

These three feedback approaches – automated, peer, and implicit – are valuable pedagogical activities, but even combined, they struggle to provide expert-level formative assessment and reliable endorsement. The questions surrounding scaling expert feedback have never been about its value, but rather its practicality. Individual expert assessment is expensive. Some efforts have focused on scaling expert-level feedback by combining it with peer review [12] or seeding peer review to infer peer grader validity [17], but these still struggle with the fundamental benefit of expert feedback: that experts have greater domain knowledge, experience, and ability to provide individual feedback. For programs that must maintain accreditation (like this university) or a guarantee (like this company), scaling expert-level review remains highly desirable.

STUDY 1: A UNIVERSITY MASTER'S PROGRAM

The first case study here presents an online Master’s program at a major university. The program is scalable, affordable, entirely online, and fully accredited. Its attempts to provide feedback at scale run into the challenges above: classes involve work that is too open-ended for automated grading alone, but too complex and high-stakes to rely on peer evaluations for grades and formative assessment. Expert-level feedback is needed, but how does it scale?

Methodology

The study began by synthesizing participant observation of the program and interviews with the program’s instructors, administrators, and TAs. First, the participant observation consisted of retrospective analysis of three years of email exchanges on the challenges to scale and their solutions. This analysis traced a narrative from an initial emphasis on automated and peer grading to the ultimate redesign of the process of recruiting and hiring TAs at scale.

Second, interviews were conducted with the head TAs (13 individuals), and some professors (six individuals) via teleconference about the administration of each class and its reliance on TAs. Most of the head TAs were on-campus PhD students; thus, the level of expertise of the interview subjects was considerably higher than that of in-progress Master’s students.

These interviews pointed to a need for continued involvement from traditional TAs. At the time of these interviews, online TAs had been hired, but sparingly. Based on the results of these interviews, we delivered a survey of the students applying to give feedback in the program, touching on their demographics, motivations, and experiences teaching in the program. This study covers the results of these interviews and surveys.

Program Background

The university's online Master's program launched in January 2014 with around 350 students, and has since grown to over 4,000 students. Three classes have graduated the program, and the average student should complete the program in three years. It is the largest such program in the United States, and has been projected to increase the annual output of Master's degrees in Computer Science by 8% [7].

Distinguishing Factors of the Online Program

While online Master's programs, and distance learning programs more generally, have been around for many years, this university's program differentiates itself in at least five significant ways [13]:

- Affordability. The entire degree costs around $7,000 [4], compared to $45,000 for out-of-state students in the residential program. Other online programs have tuitions comparable to residential tuition, e.g. Stanford's program costs approximately $63,000.
- Accessibility. Online programs naturally resolve issues of geographic accessibility, but this online program further resolves temporal accessibility by requiring no fully synchronous activities.
- Inclusivity. By removing the barriers of physical classrooms, the program can let in a greater percentage of applicants, providing opportunities to diverse groups of students who would not likely be selected with tighter space constraints.
- Custom-Built. Many programs use technology to copy processes used in person (like broadcasting live lectures), but this program constructed its courses from scratch to take advantage of the online medium.
- Accreditation. While the above factors apply to any MOOC, the key factor differentiating this online program its accreditation not only as a full Master's degree, but also as an identical degree to the residential program. There is no 'online' qualifier; the degree is the same as the one awarded on campus. This is bolstered by analysis showing that online students match or exceed residential students' performance on identical assignments [8].

These features create a unique pressure on this program: the low tuition creates a scarcity of resources with which to hire experts, but the accreditation demands a similar education and endorsement of graduates as the residential program.

Student Demographics: Anticipated Challenges to Scale

Traditionally, residential programs leverage TAs hired from the student body to support grading and feedback. To recreate the education and endorsement of residential graduates, using this same mechanism for expert feedback presents potential. Thus, to understand the challenges to scale, the demographics of the online students compared to the residential students are important. To evaluate this, some instructors begin their semesters with demographic surveys of their online and residential students.

Unsurprisingly, these surveys show dramatic differences between online and residential students. Online students are significantly older, with a median age of 38, compared to 23 among residential students. 29% of online students had previously obtained a Master's and 8% had previously obtained doctoral degrees, compared to 6% and 0%, respectively, among the residential students. Most notably, online students are far more likely to be employed: 90% reported full-time employment, compared to 5% for residential students. Finally, anecdotally, online students are far more likely to mention having children at home.

Based on these demographics and the listed tuition information, the strong hypothesis was that the online program could not scale by relying on online students the same way residential programs relied on residential students. First, the monetary incentive for online students to work as TAs is dramatically lower because residential students receive a full tuition waiver on top of their stipend. Second, the relative monetary incentive for online students is lower given their greater rate of full-time employment and existing career success. Third, the greater rate of employment and the observation that online students are more likely to support families raise the number of competing obligations; whereas residential TAs typically only balance taking classes with working as TAs, online students would also balance their family and work lives.

	Online Summer '15	On-Campus Fall '14
Median age	38	23
% previously obtaining a Master's degree	29%	6%
% previously obtaining a Doctoral degree	8%	0%
% working full-time	90%	5%
Estimated tuition per class	$510	$3,450
Estimated tuition per semester	$510	$13,800

Table 1. Demographic and tuition differences between online students in the Summer 2015 class and on-campus students in the Fall 2014 class. Online students take an average of one class per semester.

Given these observations, the assumption was that online students could not be relied upon to supply the necessary expert-level feedback; they have too many competing obligations and the financial incentive is too small. An alternative would be to hire residential students as TAs for online classes. However, it would take the entire tuitions of almost 40 online students to pay for the salary and tuition waiver of a single residential TA, and the residential program alone nearly exhausts the available TAs.

Realities of Scaling

This online program relied on residential students for TAs through the first year of the program, when it grew to around 2,000 students. This presented scaling challenges which threatened to escalate entering the summer of 2015 when most residential students leave for internships, leaving an even smaller body from which to draw.

Resolving Anticipated Challenges to Scale

Entering the summer 2015 semester, one class for the first time actively solicited teaching assistants from the online student body (two online teaching assistants had been hired for spring 2015 after they themselves expressed interest). The hypothesis remained that online students would not work as TAs, but the hope was that even a couple additional applicants would help alleviate the load.

Instead, 57 online students applied to work as teaching assistants for the class, enough to potentially support 3,000 student enrollments in that class alone. 14% of all online students that had completed the class applied. 400 students enrolled, and 10 of the 57 teaching assistants were hired; not only were there sufficiently many applicants to scale the program, but that first semester of soliciting applicants from the online students, 80% were turned away because there were not enough positions. In the year since that experimental semester, hiring for this online program has shifted almost exclusively to online students, allowing the program to continue growing by 500 students per semester.

Pedagogical Benefits of Online Teaching Assistants

The presence of the actual numbers necessary to scale the program while maintaining expert feedback is only one part of the picture. Anecdotally, nearly every professor in the program has commented on the superior performance of online teaching assistants relative to residential ones. Through conversations around these students, this trend is likely attributed to two factors. First, as online students themselves, online teaching assistants are better positioned to understand the needs of their classmates.

Second and more remarkably, many of the online teaching assistants not only have significant professional experience, but specifically have experience in the classes they are assisting. A point has been raised that graduate teaching assistants are not “experts”, but the prevalence of true experts among this body of teaching assistants is notable. The teaching assistants for the program's Educational Technology class, for example, have included an executive at a textbook publisher, two data analysts from EdTech companies, two college instructors at other schools, and a high school AP Computer Science teacher. A newly launched Human-Computer Interaction class was initially staffed by professional user experience designers and researchers from within the student body.

The expertise of these TAs far exceeds that of traditional teaching assistants. Even among more traditional former students, the number of applicants also allows professors to specifically choose the most highly-qualified former students to work as TAs in future offerings. In this way, scaling expert feedback has taken on two forms: maintaining the amount of expert feedback while increasing the size of the program, and increasing the expertise of the expert feedback through the scale and audience of the program.

Motivations of Online Teaching Assistants

The initial hypothesis was that online students would not work as teaching assistants due to low financial incentive and the more significant number of competing obligations. The justification of this hypothesis was true, but the conclusion was false; online students have applied to work as teaching assistants in droves. There are three possible explanations for this: either (a) the online applicants are coming from the subset of students most similar to residential students, and the program size is sufficient for there to be many of them; (b) we were wrong about what motivates teaching assistants in general; or (c) online teaching assistants are motivated by different factors than residential teaching assistants.

Study Methodology

To investigate this, we performed a study of the applicants and teaching assistants who either applied to work or worked within the online program in 2016. A survey was administered asking a set of questions of all applicants, with a follow-up section asking questions specifically of those that were hired. Students were solicited to participate through the list of original applications to work in the program and the list of students that were ultimately hired. The survey was sent out twice: immediately following the combined application period for Summer and Fall 2016, and immediately following the application period for Spring 2017.

	Online Applicants	**Residential Applicants**
Age		
<25 years old	11%	65%
25 – 34 years old	48%	35%
35+ years old	40%	0%
First Language		
English	81%	49%
Other	19%	51%
Employment Status		
Employed full-time	86%	8%
Employed part-time	7%	13%
Not otherwise employed	7%	79%
Employment Area		
Tech Sector	78%	14%
Non-Tech Sector	14%	7%
Not Employed	7%	79%
# of Children at Home		
0	63%	96%
1-2	30%	2%
3+	7%	2%
Highest Prior Education		
Bachelor's	80%	91%
Master's	14%	9%
Doctoral	6%	0%
Prior Experience		
>5 Years Programming Experience	72%	59%
>2 Years Teaching Experience	46%	22%

Table 2. Demographic of applicants to work as teaching assistants in the online program. Employment status excludes employment as a teaching assistant.

234 applicants completed the survey: 136 during the first distribution, 97 during the second. 161 online students and 72 residential students completed the survey. Among these, 13 of the 72 residential students who replied were hired, while 54 of the 161 online students who replied were hired.

The survey was broken into three parts: a demographic and background survey, a motivations survey, and an experience survey. All students completed the first two portions; only those that were hired completed the third.

Demographic Comparison

To begin with, the survey evaluated the demographics the applicants for teaching assistant positions to check if the observed demographics of the online program carried over into the demographics of the applicants for teaching assistant positions more specifically. The full results of this survey are shown in Table 2.

The results of the demographic portion of the survey confirm that the general trends observed in the online body of students carry over to the applicants for teaching assistant positions. Both the raw values and the comparisons to the residential applicants mirror the comparisons between the general student bodies. This confirms that it is *not* the case that the online applicants are the subset of online students that is similar to the residential applicants.

Motivation Comparison

In the second section of the survey, applicants were asked to select their single primary motivation and multiple secondary motivations. 12 options were offered, as well as a free-response box. The 12 options were inspired by the literature on learner motivation in online courses [16], and were further developed through a pilot survey asking for purely open-ended responses. These responses were summarized and coded into the 12 provided to students. Options were displayed in random order. Table 3 reports the results of these questions; the first column 'Pri.' within each group is the percent of students within that group selecting that motivation as their single primary motivation, while the italicized second column 'Sec.' within each group is the percent of students within that group selecting that motivation as their primary or one of their secondary motivations. The motivations are grouped into three general categories: Extrinsic, Intrinsic, and Altruistic.

As noted, a free-response "other" option was also supplied. No responses to this box for primary motivations fell outside the 12 response categories provided; for secondary motivations, the free response replies that fell outside the 12 provided categories were: exploring academia (3 responses), exploring new challenges (1 response), exploring a teaching career (1 response), and general curiosity about the program's inner workings (1 response).Given the lesser financial incentive available to online teaching assistants, the shift in motivations is not surprising: with a lesser financial incentive, one would expect fewer applicants to be primarily motivated by the financial incentive.

	Online Applicants		Residential Applicants	
	Pri.	Sec.	Pri.	Sec.
To obtain a tuition waiver	0%	*4%*	53%	*90%*
To obtain the salary or stipend	9%	*34%*	1%	*49%*
Any Extrinsic	**9%**	***36%***	**54%**	***96%***
To improve my resume	8%	*47%*	3%	*49%*
To improve my teaching ability	11%	*59%*	4%	*53%*
To network with faculty	11%	*53%*	11%	*49%*
To network with classmates	4%	*37%*	1%	*22%*
To learn the material	19%	*64%*	10%	*56%*
Any Intrinsic	**53%**	***93%***	**29%**	***97%***
To help my classmates	8%	*50%*	11%	*46%*
To help the instructors	4%	*39%*	1%	*32%*
To improve the class	6%	*46%*	1%	*38%*
To use my professional experience	4%	*38%*	1%	*38%*
To help the online program	16%	*70%*	1%	*17%*
Any Altruistic	**38%**	***90%***	**17%**	***76%***

Table 3. Motivations of applicants to work as teaching assistants in the online program. The 'Pri.' column designates applicants' primary motivations, and the 'Sec.' column designates applicants' secondary motivations.

What *is* remarkable is that the number of applications from this body so dramatically increased *despite* the lacking financial incentive, and these data provides an explanation of why. Online teaching assistants were far more likely than residential teaching assistants to be primarily motivated intrinsically (53% to 29%) or altruistically (38% to 17%). While both groups shared near-universal rates of some intrinsic motivators, online teaching assistants were far less likely to possess an extrinsic motivator (36% to 96%) and notably more likely to possess an altruistic motivator (90% to 76%).

More specifically, these data also demonstrate the personal ownership these online students feel over the program, given the high incidence of applicants specifying 'To help the online program' as their primary (16%) or secondary (70%) motivation. These data also suggest online applicants use working as teaching assistants to break some of the isolation experience by online students, as indicated by the relatively high percentage noting 'To network with classmates' as a secondary motivation (37%) compared to residential applicants (22%).

Given these observations, we conclude that it is the case that online students retain different motivations for applying to work as teaching assistants than residential students. This finding on its own is unsurprising; however, when taken in combination with the significant number of applicants received from this audience, it provides guidance for ways to recruit and incentivize online students to help programs with high demands for expert feedback and endorsement.

Experience Comparison

As noted previously, 13 of the 72 residential respondents and 54 of the 161 online respondents to the survive were then hired to work as teaching assistants in the program. The final portion of the survey focused on their experience as teaching assistants. In this phase of the survey, there was relatively little difference between online and residential teaching assistants in terms of workload and responsibilities. Both primarily focused on manually grading assignments, projects, and exams, echoing the premise of this paper that the primary effect of this initiative is scaling expert feedback. Interestingly, residential teaching assistants reported a median workload per week of 12 hours, while online teaching assistants reported a median workload of 10 hours; this is notable because online teaching assistants are paid hourly, while residential teaching assistants are salaried with an assumption of 20 hours per week. If this ratio generalizes, this would cut the cost per teaching assistant even further.

As noted, anecdotally, most professors have noted superior engagement, ownership, and performance from online teaching assistants, although there are many exceptions. Generally, differing responsibilities make formal, direct comparisons difficult. However, one comparison corroborates these anecdotes. One class employed an all-residential teaching team one semester, and an all-online teaching team the next; the two teaching teams graded the same assignments, a set of six 1000-word essays. The all-online team gave an average of 130 words of feedback per assignment, while the residential team gave an average of 22 words per assignment. The least prolific online teaching assistant gave an average of twice as much feedback as the most prolific residential teaching assistant.

Conclusions of Study 1

This university's new online Master's program has, at a minimum, preserved the quantity of expert feedback while scaling an online Master's program to 4,000 students. Strong arguments can be made as well that the program has increased the volume and quality of that feedback in addition to merely preserving it with the growing program.

This growth has been made possible by an unanticipated audience of online students interested in giving back to the program as teaching assistants. The fact that these students are largely intrinsically or altruistically motivated is not on its own overly remarkable; many MOOC programs, such as Coursera's mentor program [26], have similarly noted similar student bodies. What makes this development in this online program notable is that this audience can actually support a fully accredited, rigorous, prestigious program with high standards for success. The program has relied on this audience to create a groundbreaking program [4] that preserves learning outcomes [8] and the student experience [13] while dramatically increasing size, affordability, and accessibility.

It is worth noting, as this is a case study, that there are questions as to the generalizability of these observations. As noted by the survey results, the vast majority of online applicants apply in part to help the program; this echoes a sense of ownership over the program. We hypothesize this comes from (a) the desire to participate with a revolutionary program [4], and, (b) a sense of gratitude toward an opportunity most students would not have otherwise had [13]. These motivations may not generalize in the same way to non-accredited programs, nor are they guaranteed to persist as programs like these become more common. Interestingly, however, this may also suggest that the accreditation and prestige of the program both dictates and resolves the need for increased expert-level feedback, as participating in that prestige may be part of students' motivations to help the program.

STUDY 2: AN INDUSTRY MICROCREDENTIAL

The second case study focuses on a for-profit online education provider that supplies project-based microcredential programs in fields like machine learning, virtual reality, and web development. Its programs generally offer students the opportunity to go at their own pace, paying a monthly subscription cost if they remain in the program. Most notably, the microcredential programs are entirely project-based, and typically feature open-ended, partially student-defined projects.

Unlike the university, the company's microcredential programs do not have the pressures of accreditation demanding the presence of expert-level feedback and endorsement. However, a similar pressure arises in a different form: the company offers a job guarantee for graduates from their programs in machine learning, web development, data analysis, and mobile development. With the promise that any learner completing the program would receive a job or their money would be refunded, the pressure to possess both authentic, project-based assessments and reliable endorsements of learner ability rose considerably. However, assessing and teaching through authentic, open-ended projects necessitated expert feedback.

Methodology

This case study emphasizes the mechanisms that give rise to this system of feedback and the results of the system in action. As such, the primary sources of information for this case study are the design documents of the system architecture and the data automatically generated during its regular use. In addition to evaluation of the design documents, three interviews were conducted via teleconference (notes taken by hand by the interviewer) with individuals involved in the system in some capacity: an engineer working on the system, a content developer creating content for the system, and a reviewer evaluating work through the system. These interviews focused on the goals of the system in achieving rapid, scalable, high-quality feedback. Although these interviews provided valuable backdrop, the major takeaways below are the design of the system and the measurable results achieved.

Project Reviewer Infrastructure

Rather than rely on the traditional model of hiring on-staff teaching assistants to evaluate and give feedback on projects, the company instead developed a network model for providing expert reviews. This system leverages a distributed network of freelance project reviewers paid per project that they evaluated.

Training Project Reviewers

Project reviewers were drawn from three audiences: professionals working in the field, exemplary microcredential graduates, and course developers and managers that helped produce the programs in the first place. Once identified, new reviewers go through a training process. They participate in a general course on reviewing and giving feedback on projects, followed by some additional material specific to the field in which they will be reviewing, highlighting the type of feedback to give and the misconceptions to anticipate.

After the training course, prospective project reviewers are provided multiple sample projects to evaluate, and their evaluations are provided to a set of super-reviewers. These super-reviewers, themselves experienced project reviewers, give the prospective reviewers feedback on the degree to which the evaluation matched the evaluations expected (in both result and feedback), as well as what the strengths and weaknesses were of the feedback provided.

In this way, prospective project reviewers effectively participate as students in the review system, where their "projects" are reviews of others' projects. This plays two roles: not only does it provide the prospective reviewers with feedback to help ensure they align, both in quality and in conclusion, with the other reviewers, but it also allows them to experience the system from the student perspective.

Project Reviewing in Action

When a student in a microcredential program submits a project, the project becomes available on the dashboards for any reviewers approved to review that project. This process is instantaneous, a notable contrast to the traditional deadline- and batch-based grading methods used by many areas of higher education. This is afforded by the microcredential programs' self-paced nature.

Once a project appears on the dashboard to be reviewed, any certified reviewer can claim it. After claiming the project for review, the reviewer is supplied with the project itself (typically some source code with a written report or documentation), notes from the student, and a history of the project submission. In this way, reviewers can look at the historical submissions the student has made to evaluate the new submission in context. Reviewers were observed using this to comment on student progress specifically in the context of the progress that was already observed.

Projects are evaluated on a pass/fail basis across multiple criteria. For a project to pass, it must pass each individual criterion; if it does not, the reviewer provides feedback on what revisions will be necessary to meet the project's expectations.

Project Review Results

Four metrics are available for evaluating the success of this novel project review process: scale, turnaround time, learner satisfaction, and reviewer earnings.

Evaluation of the Project Reviewer System

First, the primary motivation of this system is to scale up the number of projects that can be evaluated. At time of writing, the project reviewer system described here is processing approximately 650 projects per day. Thus, in terms of scale, this system appears successful, reviewing an average of a project every two minutes.

Second, research shows that the speed at which feedback is received is a significant determinate of learning outcomes [3, 18]. This has presented a challenge for traditional deadline-based models of education because evaluation tends to wait until the deadline has passed, regardless of how far in advance of the deadline an individual submits their assignment. After a deadline has passed, it would typically take at least a day for the batch of feedback to be processed, if not far higher. Given that this project reviewer infrastructure is always-available and the microcredential programs themselves lack deadlines, these constraints are absent. Thus, a second metric for evaluating the system's effectiveness turns up similarly positive results: the median turnaround time for a project review is 92 minutes. Under this mechanism, a student could conceivably complete multiple iterations of a project in a single day, complete with expert feedback. Further research will evaluate the prevalence of this rapid revision.

Third, although student satisfaction does not guarantee positive learning outcomes, it nonetheless provides a glimpse into the learners' impressions of the system. Learners within this system are asked to meta-review each review they receive out of 5 points. The average rating assigned to reviews that are received is 4.9. This infrastructure also allows auditing of project reviewers if multiple subpar reviews are received.

Thus, this project review system generates many reviews rapidly that satisfy learners. The university covered in the first case study found similar results, but without a heavy financial incentive. Although no survey has yet been performed on the project reviewers within this company's system, we hypothesize the motivations in this case *are* extrinsic. First, the intrinsic or altruistic incentives present in the university's program do not appear to generalize here: there are no professors with whom to network, the project reviewers do not heavily collaborate with one another, and we do not see the same signs of a sense of ownership over the program among the company's project reviewers as we see in the university's teaching assistants. Second, the financial incentive in this company's project reviewer system is very significant; the top-paid project reviewer currently earns over $4,500 per week. This heavy earning potential creates a strong incentive to maintain high quality, given the threat of competition.

Pedagogical Benefits

It is worth noting that the metrics described above do not directly capture the learners' learning processes. That learners are satisfied with their reviews does not guarantee positive learning outcomes, and nor does rapid feedback guarantee quality feedback. The metrics demonstrate successful accomplishment of scale, but these metrics do not directly capture learning outcomes. For this, we analyzed more theoretically the pedagogy that results from this review system. Follow-up analyses will examine the prevalence of revision and the workflows involved therein.

In practice, several pedagogical benefits were observed emerging out of this system. First, research has shown that the speed with which feedback is received has a strong connection to learning; rapid feedback leads to better, more rapid learning [3, 18]. This project reviewing framework facilitates rapid feedback, as demonstrated by the 90-minute median turnaround time between submission and receipt of evaluation. That rapid feedback cycle is further enhanced by the second pedagogical benefit, the opportunity for revision. Rather than receiving a grade and moving on as is common in many learning environments, learners participating with this project review system can iterate on and improve their solutions. The strength of the rapid feedback cycle becomes even more pertinent because the feedback is directly applied to a revision of the same work.

These pedagogical benefits also enhance the positive effect seen from other elements of the programs' structures. For example, the self-paced nature of these programs removes the high-stakes grades, allowing learners the opportunity to iterate more naturally without the heavy extrinsic pressure of grades and deadlines. Additionally, and especially pertinent for the fields these programs address, the notion of iterating over a project is authentic to the domain. Learners thus learn the metacognitive knowledge involved in evaluation and revision that is similarly part of the domain skillset they are developing.

Conclusions of Study 2

Where the university described in the first study scaled up its enrollment while preserving the involvement of experts by relying on a shift in motivations to maintain a traditional system, the company here in the second study instead opts for a radically different system. Rather than traditional teaching teams under the direction of a single faculty member, this company uses a distributed network of on-demand project reviewers. This system has proved successful in dramatically increasing the number of projects evaluated daily, while introducing additional pedagogical benefits as well.

As with the first study, however, there are questions as to the generalizability of this second study. First, part of the flexibility to design a system like this comes from the company's non-academic nature; it is not governed by details of FERPA and other regulations, allowing more freedom to select its evaluators. Secondly, this type of system does rely on the presence of an existing body of experts capable of evaluating projects. Moreover, it relies on the ability to provide a sufficient financial incentive to such a body to have them contribute to reviewing projects. These challenges may threaten the ability of a system like the one described here to generalize to academic environments or other content domains.

CONCLUSION

These case studies do not aim to argue that expert feedback is "better" than the scalable feedback mechanisms used in other online programs, like automated feedback, peer feedback, and implicit feedback. Indeed, the academic program reviewed in this paper leverages all four types of feedback in its offering, and removing any one of these four would have detrimental effects. It is also worth noting that the importance of expert-level feedback is derived from the demands of public perception and documented weaknesses in alternative feedback methods; however, this analysis does not attempt to present evidence of superior learning gains from expert-level feedback compared to other methods. If it is the case that other feedback methods may present the same learning gains, then expert-level feedback may itself be unnecessary.

If expert-level feedback has value, however, there are three primary takeaways from these case studies. First, whether in accredited academic programs or in vocational industry programs, there remain pressures to supply expert-level feedback and reliable endorsement of graduates' abilities. Solely relying on automated, peer, or implicit feedback presents challenges to the pedagogy and perception of these programs. Even setting aside most common criticisms of these methods, these methods all struggle with providing expert-level, targeted, individualized feedback on open-ended projects at scale. This struggle presents a problem for programs that experience significant pressure to retain open-ended, authentic projects, expert-level feedback, and reliable endorsement.

To facilitate scale, many programs have dropped expert-level feedback altogether, and there exists a strong argument that the additional accessibility and affordability justify that loss. However, this paper shows two similar methods that have been successful at increasing scale while retaining expert-level feedback. The second takeaway of these case studies, then, is that it is possible – at least for now – to rely on the documented unique motivations of online students in these innovative programs to maintain or increase the amount of expert-level feedback while retaining a traditional structure, even as the financial incentives become several times less lucrative. The third takeaway is that it is also possible to develop an entirely alternative model, based not around traditional teaching assistants but rather around a distributed network of reviewers, that radically increases the rate of feedback while preserving the ability to offer authentic projects in a highly affordable program. Both these methods are worth exploring in other domains and programs.

Taken all together, this analysis argues that there remains a need to have expert evaluation in high-stakes (accredited or otherwise guaranteed) programs. We have documented two options for this: One, maintaining the traditional model dictates that the financial incentives for traditional teaching assistants will be lower, but the shifting motivations of online students are sufficient to provide a strong pool of experts nonetheless. Two, throwing out the traditional model, it is possible to leverage a more agile distributed workforce to provide rapid expert feedback on open-ended projects while preserving affordability. These two methods cover some of the space surrounding scaling expert feedback; other methods that may be explored include relying strictly on a volunteer model (such as Coursera's mentorship program [26]). Given the role of altruistic and implicit motivations in scaling expert feedback, scale through a purely volunteer model may be possible if the necessary accountability can be duplicated while preserving these motivations.

REFERENCES

1. Aleven, V., Mclaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education*, *19*(2), 105-154.
2. Boud, D., Cohen, R., & Sampson, J. (2001). *Peer Learning in Higher Education: Learning from & with Each Other*. Psychology Press.
3. Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245-281.
4. Carey, K. (2016, October 5). Georgia Tech's $7,000 Online Master's Degree Could Start a Revolution. *New York Times*, pp. A15.
5. Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287-322.
6. Geigle, C., Zhai, C., & Ferguson, D. C. (2016, April). An Exploration of Automated Grading of Complex Assignments. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (pp. 351-360). ACM.
7. Goodman, J., Melkers, J., & Pallais, A. (2016). Can Online Delivery Increase Access to Education? HKS Faculty Research Working Paper Series RWP16-035.
8. Goel, A. & Joyner, D. A. (2016a). An Experiment in Teaching Cognitive Systems Online. In Haynes, D.

(Ed*.) International Journal for Scholarship of Technology-Enhanced Learning 1*(1).

9. Goel, A. & Joyner, D. (2016b). *Formative Assessment and Implicit Feedback in Online Learning*. Presentation, Learning with MOOCs III.

10. Hext, J. B., & Winings, J. W. (1969). An automatic grading scheme for simple programming exercises. *Communications of the ACM*, *12*(5), 272-275.

11. Jensen, J. C., Lee, E. A., & Seshia, S. A. (2013, April). Virtualizing cyber-physical systems: Bringing CPS to online education. In *Proc. First Workshop on CPS Education (CPS-Ed)*.

12. Joyner, D. A., Ashby, W., Irish, L., Lam, Y., Langston, J., Lupiani, I., ... & Bruckman, A. (2016, April). Graders as Meta-Reviewers: Simultaneously Scaling and Improving Expert Evaluation for Large Online Classrooms. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (pp. 399-408). ACM.

13. Joyner, D. A., Goel, A., & Isbell, C. (2016). The Unexpected Pedagogical Benefits of Making Higher Education Accessible. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*. ACM.

14. Kendall, K., & Schussler, E. (2012). Does instructor type matter? Undergraduate student perception of graduate teaching assistants and professors. *CBE-Life Sciences Education*, *11*(2), 187-199.

15. Kirschner, P., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86.

16. Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *22*(2), 6.

17. Kolhe, P., Littman, M. L., & Isbell, C. L. (2016, April). Peer Reviewing Short Answers using Comparative Judgement. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (pp. 241-244). ACM.

18. Kulkarni, C., Bernstein, M. S., & Klemmer, S. (2015). PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. In *Proceedings from The Second ACM Conference on Learning @ Scale*. ACM.

19. Lu, Y., Warren, J., Jermaine, C., Chaudhuri, S., & Rixner, S. (2015, May). Grading the Graders: Motivating Peer Graders in a MOOC. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 680-690). ACM.

20. Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30-43.

21. Murray, T., Blessing, S., & Ainsworth, S. (2003). Authoring tools for advanced technology learning environments: Toward cost-effective adaptive, interactive and intelligent educational software. Springer Science & Business Media.

22. Polson, M. C., & Richardson, J. J. (2013). *Foundations of intelligent tutoring systems*. Psychology Press.

23. Raman, K., & Joachims, T. (2015). Bayesian Ordinal Peer Grading. In *Proceedings from The Second ACM Conference on Learning @ Scale*. ACM. 149-156.

24. Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529-556.

25. Salzmann, C., Gillet, D., & Piguet, Y. (2016, February). MOOLs for MOOCs: A first edX scalable implementation. In *2016 13th International Conference on Remote Engineering and Virtual Instrumentation (REV)* (pp. 246-251). IEEE.

26. Saraf, K. & Smith, C. (2016, October 5). "Coursera's superheroes: Meet the Mentor team." *Coursera Blog.* Retrieved from coursera.tumblr.com/post/151389966612/

27. Srikant, S., & Aggarwal, V. (2014, August). A system to grade computer programming skills using machine learning. In *Proceedings of the 20th ACM International Conference on Knowledge Discovery and Data Mining* (pp. 1887-1896). ACM.

28. VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, *16*(3), 227-265.

29. VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.

30. Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*(2), 155-174.

31. Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, *20*(4), 270-279.

32. Vogelsang, T., & Ruppertz, L. (2015, March). On the validity of peer grading and a cloud teaching assistant system. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 41-50). ACM.

33. Yaron, D., Karabinos, M., Lange, D., Greeno, J. G., & Leinhardt, G. (2010). The ChemCollective—virtual labs for introductory chemistry courses. *Science*, *328*(5978), 584-585.