Informal Learning Communities:
The Other Massive Open Online 'C'

Will Hudgins, Michael Lynch, Ash Schmal, Harsh Sikka, Michael Swenson, and David A. Joyner

College of Computing, Georgia Institute of Technology

Atlanta, GA, USA

{wkhudgins, mikelynch, kschmal3, harshsikka, swenson, david.joyner}@gatech.edu

ABSTRACT

While the literature on learning at scale has largely focused on MOOCs, online degree programs, and AI techniques for supporting scalable learning experiences, informal learning communities have been relatively underrepresented. Nonetheless, these massive open online learning communities regularly draw far more engaged users than the typical MOOC. Their informal structure, however, makes them significantly more difficult to study. In this work, we take a first step toward attempting to understand these communities specifically from the perspective of scale. Taking a sample of 62 such communities, we develop a tagging system for understanding the specific features and how they relate to scale. For example, just as a MOOC cannot manually grade every assignment, so also an informal learning community cannot approve every contribution; and just as MOOCs therefore employ autograding, informal learning communities employ crowd-sourced moderation or platform-driven enforcement. Using these tags, we then select several communities for deeper case studies. We also use these tags to make sense of learning-based subreddits from the popular community site Reddit, which offers an API for programmatic analysis. Based on these techniques, we offer findings about the performance of informal learning communities at scale and issue a call to include these environments more fully in future research on learning at scale.

Author Keywords

Informal learning; informal learning communities; Reddit; case studies; online learning

CSS Concepts

• Social and professional topics~Informal
tion   • Applied computing~Education   • Applied computing~Collaborative learning   • Applied computing~Distance learning   • Applied computing~E-learning

INTRODUCTION

The literature in Learning @ Scale has largely been dominated by two themes with considerable overlap: MOOCs and higher education [10]. Much of the research has focused explicitly on these domains, while other portions of the community are dedicated to specific tools and technologies that, while theoretically somewhat generally applicable, find their main usefulness in these areas. Abstracting out, there has been research also on other formal learning environments, such as K-12 initiatives and ESL classes.

Informal online learning communities, by comparison, are relatively underrepresented. This is likely due in part to the inherent overlap between university researchers and university course development, but also likely derives from the difficulty in evaluating these communities. It is inherently difficult to know who the learners are, including their demographics, goals, and outcomes. There exists no start-of-course "survey", no pre-test and post-test on learning outcomes, and no formal enrollment or dropout mechanism.

And yet, these communities are potentially more active and impactful than many of the MOOCs and online degree programs that have received significant research attention. As our analysis will who, the Language Learning subreddit (reddit.com/r/languagelearning) counts almost 240,000 subscribers, of which 6.55% are considered "engaged". Thus, at over 15,000 engaged users, the community has over twice the median number of active members as the average MOOC in a report compiled by Harvard and MIT in 2017 [3]. These communities therefore present a legitimate alternative to the more eye-catching MOOCs. However, as noted above, they are far more difficult to study. There have been efforts, of course, such as the ground-breaking study on Fanfiction.net by Aragon & Davis [2], but these existing research efforts have largely focused on informal learning communities from a more individualistic perspective.

The purpose of this research is to understand informal online learning communities from the perspective of scale; what allows these communities to scale? Specifically, this research addresses the following questions: what are the characteristics of informal learning communities that achieve scale? What are the major different communication and dissemination methods used by informal learning communities to support learning? How can informal learning communities be categorized? What characteristic(s) should typologies be

based on? How can future researchers expand on the research accomplished in this project?

This project attempts to answer these questions by establishing essential categories for informal learning, delineating them according to the nature of the community itself, the content that is shared there and how it is produced, the characteristics of the platform and, where applicable, the sub-communities it hosts. As this research attempts to evaluate these communities from the perspective of scale, we treat the communities themselves as the target of analysis rather than individuals in the communities; thus, characteristics of members of these communities were excluded from our analysis. These community types are further defined by a system of tags that we developed.

Defining Terms
The distinction between formal and informal learning can be contentious, and researchers have proposed a variety of (at times, conflicting) definitions [5]. In general, there appear to be two main modes of thought, with some contending that formal and informal thinking are two distinct modes of learning [7][13], while others argue that formal and informal merely describe different attributes of learning, and that nearly all learning environments comprise some mixture of both formal and informal elements [12].

Within these two groups, definitions still vary. The distinction between formal and informal may depend heavily upon the context of where the learning takes place; for example, the definition for formal and informal learning at the workplace may differ by necessity from the definition for formal and informal learning in an academic institution [4][8][15]. Moreover, previous researchers often define formal and informal learning by describing both environmental attributes and learner attributes [7].

In recognition of the difference context can make in a learning environment, we focus on defining informal and formal learning only as they pertain to online learning environments. In order to avoid making assumptions about individuals' motivations to participate in these communities and the unnecessary exclusion of certain communities based on these assumptions, we define informal learning based only on the attributes of the environment, not the learners. We propose that informal online learning environments possess the following characteristics:

- Learning is democratic. While learning may be facilitated by a small group of individuals in some cases, informal online learning environments focus on shared knowledge and the development of learning materials that are provided by a large number of participants.

- Learner knowledge is not evaluated in any standardized manner that is *built in* to the environment. An informal community may arise dedicated to studying for a specific standardized exam or evaluation, but the exam itself must not be administered as part and parcel to the community for the community to be considered informal.

- Learner participation is not authenticated with any form of certification or diploma for external purposes, although users may receive in-platform recognition from their peers (such as reputation points).

Notably, these definitions still leave plenty of room for ambiguity. For example, in our analysis, we consider the sub-reddit /r/volleyball to be a "learning" subreddit; while the organization is also for general networking among volleyball aficionados, we noted that both in the rules and in the posting patterns, users regularly share educational content, either teaching how to play the sport or teaching about the history of the sport. At a certain level, any content-sharing community could likely be interpreted as "learning"-based, as the goal is to learn about the content shared. For this reason, in our analysis we have avoided making any general claims about all informal learning communities, as this would require a more formal definition of what differentiates learning communities from other communities; instead, we offer only a way of interpreting informal learning communities, and leave further delineation to future researchers.

RELATED WORK
Several typologies have been proposed for the categorization of online courses [11]. Typically, these focus on categorizing online learning in terms of the technology used [1][6][9], the pedagogical nature of the course [20], or the relationship between teachers and students [14][18]. Riel & Polin, for instance, suggest such communities tend to be either knowledge-based, practice-based, or task-based [17]. Such categorizations provide a useful foundation on which to base further discussions of online learning. However, many of these proposed typologies are dated with respect to the rate at which the Internet evolves. While online courses were among the early recognized forms of online learning, new environments for online learning and instruction have emerged.

Additional typologies surrounding the nature of online communities more generally have been proposed. Stanoevska-Slabeva, et. al., grouped communities into one of four types: discussion-based, task and goal-oriented communities, virtual worlds, and hybrid communities [19]. The authors further identified common platform elements for each community type, with the goal being to propose a system for the ideal design of new online communities. More recently, Porter grouped online communities by "establishment" (member vs. organization led), and further, by the nature of the establishment (social or professional for member-led communities; commercial, non-profit, or government for organization-led communities) [16]. Porter notes that communities can be further described according to their purpose, place ("extent of technology mediation of interaction"), platform, population interaction structure, and profit model. Many of the characteristics described by these typologies are similar to those of our own, though they are non-specific to learning communities.

METHODOLOGY

Our methodology follows three stages: first, an effort to articulate common characteristics among informal learning communities. Second, we apply the tags that result from that effort to a larger sample of such communities. Third, we take a deeper look at five of these communities, using the tags to ensure that we are looking at communities that operate in different ways. The second and third phases of this analysis were iterative, with deeper case studies informing further revision to the tagging system; they are presented sequentially here to better summarize the results.

To begin, we performed an ad-hoc search for Informal Learning Communities using various web-based search engines with various terms designed to find communities that were focused on learning of any kind. The goal of this initial step was not to gather a representative sample, but rather to gather a varied sample of the range of informal learning communities; thus, this random sampling approach was deemed acceptable. All methods including inclusion and exclusion criteria were determined a priori. The only exclusion criterion was that they must not be private: they had to be public, although supporting free account creation was deemed acceptably public. We examined 62 total sites.

We then conducted deeper case studies on selected communities to identify defining platform and social elements for each. A member of the research group visited each online community and captured an evaluation of each site, such as how they communicated, what specific features their community offered, and information about the size of the membership and activity of the community. Following these in-depth qualitative examinations, a preliminary tagging system was developed to describe the elements identified in this search. Refined versions of these case studies are presented later in the paper.

Using this initial tagging system, we then expanded our list of communities. The tags were applied to each of the communities based on the review. Additional tags were added to the list as new features were identified in the extensive review. The final list of communities and their tags is available for review in the appendix.

In describing these communities, we chose to focus solely on characteristics that described features built into the platform itself, such as social structures within the community (standards, enforcement, social rewards, etc.), and features that described the "goals" of the learning community. We avoided any tags that might relate to learner characteristics, such as those that might describe demographics, motivation, etc.; we find these have been more thoroughly researched elsewhere, and our goal is to examine the design of the communities themselves for scale. Additionally, while previous typologies have tried to incorporate both learner characteristics and community characteristics into the same system, we feel there is value to understanding community and learner characteristics separately. While there is certainly a relationship between community and learner characteristics in online learning communities, attempting to combine these two without first looking at them separately can also lead to conflicting or otherwise confining definitions of different community features.

Many of the communities we identified were "subreddits" on the web site Reddit.com, essentially communities that exist on a common platform for supporting a wide range of other communities as well. Reddit offers a powerful public API facilitating extraction and analysis of Reddit activity (Reddit, 2020). To take advantage of this, we built a custom scraper to facilitate extracting data from the Reddit API using the PRAW Python Reddit API wrapper. For every subreddit we identified, we collected the author ID, comment ID, link ID, parent ID, submission ID, subreddit ID, subreddit name, and creation timestamp. We also determined if the comment was the original post in a thread and whether the comment was created by the original poster. We collected this data for the time range from November 15, 2018 (UTC) to November 15, 2019 (UTC).

Scope

This study focuses on surveying several online learning communities, designing several structured qualitative case studies that analyze various aspects of the communities as they pertain to informal learning, providing a general framework to categorize them, and using that approach to categorize several communities. Accompanying data includes qualitative analysis of communities along several criteria axes, as well as quantitative data outlining usage statistics, community activity, and other web available information. This analysis will provide researchers with a qualitative direction on how informal learning communities may be characterized, and the framework developed will allow researchers to evaluate typing as a methodology, extending or replacing it in the overall endeavor of understanding online informal learning. Communities were sampled based on qualitatively selecting for content, activity, and were chosen through the authors' personal experience and network. Qualitative analyses were done by the authors and involved observing communities and understanding their activities, emergent behaviors, rules and moderation based on the available information without deep, long term. participation. Tags used in building the typology were also qualitatively designed and are included to demonstrate the usefulness of categorizing informal learning communities for research purposes.

Limitations

It is important to note that this analysis does not aim to be a comprehensive summary of the distribution of online communities; our sampling mechanisms are not random enough to make claims about (for example) the number of all communities that possess each of these tags. Instead, our claims are about the existence of communities possessing these criteria, the value of applying these tags to communities to support making sense of the landscape, and the ways in which these criteria interact with the behavior of the communities. In short, these are case studies.

The exception to this is the quantitative data derived from Reddit; even here, though, the communities we selected are subject to our sampling methods. We do not make claims about the nature of all Reddit communities, but rather only pose these communities as interesting exemplars.

INFORMAL LEARNING CATEGORIZATIONS
Below is a table of the proposed "tags" that can be applied to different learning communities, along with a description of the tag. We divide the tags into several meta-categories.

Learning Activity
The learning activity is the actual loop within which the learning takes place by the users; what is the nature of the learning?

Discussion-Based Learning: Learning stems from discussion where the knowledge to be gained as fewer definitive "facts" or "correct" answers.

Creation-Based Learning: Members create the content the community is intended to consume. Content is created by individuals or content can be created by small groups. Content is often created through an iterative editing process.

Tutorial-Based Learning: Instructional content is presented via how-tos and guides; the informal learning community are learners together following and discussing those tutorials.

Interaction Format
Interaction format refers to the structures which learners use to communicate with one another, including the inferred structure of the interaction. These are not exclusive categories; interaction could be real-time and Q&A-based, for instance.

Real-Time Communication: Communication includes a chat room or instant-messaging function that allows members to communicate in near real time. (Example: Discord chat and voice communication.)

Forum-Based Communication: Communication between members takes place in a post and response style. (Example: Listservs and forums.)

Q&A-Based Communication: Communication includes a Question & Answer function where the primary post is a question to the community or instructor.

Direct (Private) Messaging: Communication includes a direct and private member-to-member function. Messages sent in this way are usually only accessible to the two members.

Rules & Incentives
Rules and incentives are the elements of a community's structure that guide interaction, either by delineating what is allowed and disallowed (rules) or by rewarding desirable behaviors (incentives). These also include enforcement structures: how are rules enforced or incentives given?

Community Standards Present: Community standards govern social etiquette within the community. These standards are typically listed in a conspicuous and clear manner

and govern the ways in which participants are expected to behave toward each other.

Community Standards Enforced by Community: Community standards are largely enforced by members of the community themselves either through peer pressure and widespread member buy-in and/or through the empowerment of members who serve as content moderators with the ability to make executive decisions about the suitability of certain interactions.

Reward System Present: Frequency and/or quality of individual member participation is rewarded and displayed through a badge or point system.

Quality Standards Present: Quality standards govern the expected attributes of any instructional content or information that is posted to the platform by its members.

Quality Standards Enforced by System: Quality standards are enforced implicitly rather than through the explicit removal or editing of deficient instructional content/information. This can be done democratically (members vote on, or rate the quality of the content), or in some cases, marked by an individual. These votes/ratings are then displayed next to the accompanying instructional content.

Location
Location refers not to geographic location, but virtual location: does the community exist on a shared, centralized site, like a particular subreddit, or is it dispersed across multiple sites, such as informal learning communities dedicated to following particular YouTube courses, discussing via Twitter hashtags, email groups, or multiple Slack or Discord organizations?

Centralized Community: Communities that are centralized in a specific "location". These can be sub-communities of a larger site (such as a "sub-reddit" or other kind of sub-forum) so long as the entirety of the community can be easily found in a singular, labeled space.

Dispersed Community: Communities that are spread out across a more generalized platform in a non-obvious and sometimes "invisible" manner. These communities do not have a designated section of their greater platform that is dedicated to collating their instructional content. Members must find each other via networking across the platform.

Contains Sub-Communities: Platforms represent a community in and of themselves while also hosting smaller sub-communities that may take a different format from that of the primary platform or community. This is distinct from platforms that contain a large number of sub-communities which all take the same format and where the primary platform is not itself a connected learning community.

Features
Features are specific tools or functions built into the structure of the community's platform.

Teacher Present: A "teacher" is a single individual or a small group of individuals with greater instructional authority than other members of the group. Communities with a teach-

er present flag may still have members that provide additional-al know-how. (Examples: WikiHow instructors.)

Meta-Community: Platforms where discussions about community rules, standards, and content moderation take place behind the scenes and are conducted by a sub-community that is often "invisible" or non-obvious from the front-facing site. Membership to the meta-community is open, though members of the meta-community are typically expected to have a higher degree of investment and commitment to the community.

Private: Members must register in order to participate actively in the community.

Tag System Present: Content can be categorized by a formal labeling system such that a member can easily find related content by searching with a specific tag.

Content

Content tags refer to the types of content that form the foundation of the informal learning community. We identify three main content tags: the artifacts the learners produce, the critiques learners produce about others' artifacts, or content that learners are navigating together.

Artifacts: Members post and share artifacts they own that have been created by them or have created with other members. (Example: DeviantArt.)

Critiques: Members may post content and receive responses specifically to provide constructive feedback to the member about their work shared in the post. (Example: Fanfiction.net comments.)

Goal

Goal tags refer to the shared general intention of the community, whether viewed individualistically ("learners want to learn programming") or collectively ("the community want to support learners learning to program").

Skill Building: Communities that are focused on helping individuals improve a specific skill, rather than on the accumulation of knowledge about a given subject (though knowledge about the subject may be gained in the process). (Examples: language learning communities.)

Focused: Focused or goal-oriented communities, where the central topic of the community is focused on a specific domain (e.g. "to learn computer programming") in contrast with some communities focused on general knowledge (e.g. "Explain Like I'm 5", a general knowledge community for explanations on any topic).

QUANTITATIVE ANALYSIS

We applied the tagging system to a larger number of informal learning communities that we identified. For those communities identified as residing on Reddit, we further made use of the powerful Reddit API to scrape and summarize these communities. This quantitative analysis provides some useful insights, as well as valuable context for the deeper case studies that follow.

Tag Frequency

We first reviewed our tagging system to understand the distribution of communities in our sample. These percentages do not summarize the distribution of all informal learning communities, but rather to provide a glimpse at the types of communities that were common within our sample.

Table 1: Tags and their frequency within our sample of informal learning communities.

Tag	% of communities with tag
Standards enforced by community	74%
Standards enforced by software/platform	74%
Quality standards present	73%
Community standards present	71%
Forum-based Communication	69%
Reward system present	60%
Quality standards enforced by system	58%
Q&A-based Communication	55%
Discussion-based learning	55%
Focused	50%
Direct (private) Messaging	48%
Artifacts	45%
Centralized community	45%
Tag system present	45%
Critiques	44%
Creation-Based Learning	35%
Skill building	32%
Meta community present	19%
Contains Sub-communities	18%
Tutorial-based learning	16%
Private	13%
Real time Communication	10%
Teacher Present	8%
Dispersed community	6%

As noted, this is not meant to be a representative sample of all informal learning communities. Nonetheless, it is interesting that the communities in our sample lean strongly toward community- and platform-enforced standards for both behavior and quality, that reward systems are commonly used in addition to these rule-based standards, and that asynchronous, forum-based communication is significantly more common than real-time communication. These trends may reflect the differential ease with which these communities can be found; a forum-based community, for example, is subject to more search engine crawling than a private chat-based community. We note, though, that those communities easier to find for research may also be easier to find for new potential participants.

Table 2: Engagement data from several learning subreddits. EpRM stands for Engagements per Registered Member. AE/EU stands for Average Engagements per Engaged User. AI/I stands for Average Initiations per Initiator.

Subreddit	Engaged Users	Exclusive Respond- ers	Initiators	Initiator- Respond- ers	Exclusive Initiators	EpRM	AE/EU	AI/I	Critique Tag	Focused Tag
answers	6.79%	62.18%	37.82%	7.16%	81.05%	38.45%	5.66	2.21	No	No
AskHistorians	1.03%	31.24%	68.76%	5.35%	92.20%	3.88%	3.76	1.76	Yes	Yes
AskReddit	0.32%	91.56%	8.44%	1.42%	83.21%	0.79%	2.47	1.35	No	No
askscience	0.09%	46.15%	53.85%	1.56%	97.10%	0.20%	2.21	1.44	Yes	Yes
DMAcademy	12.84%	70.22%	29.78%	12.53%	57.92%	74.30%	5.79	1.53	Yes	Yes
explainlikeimfive	0.21%	80.27%	19.73%	1.23%	93.79%	0.55%	2.60	1.47	No	No
languagelearning	6.55%	70.57%	29.43%	9.24%	68.57%	37.74%	5.76	1.87	Yes	Yes
learnmachinelearning	5.22%	59.53%	40.47%	8.43%	79.13%	16.89%	3.24	2.25	Yes	Yes
learnprogramming	1.61%	62.29%	37.71%	5.54%	85.30%	7.20%	4.48	1.51	Yes	Yes
lifelonglearning	0.47%	42.42%	57.58%	0.00%	94.74%	1.04%	2.21	2.50	Yes	No
LifeProTips	0.47%	92.26%	7.74%	1.39%	82.00%	1.12%	2.38	1.42	No	No
todayilearned	0.75%	96.21%	3.79%	0.90%	76.30%	2.26%	3.01	1.76	No	No
UniversityofReddit	0.06%	84.91%	15.09%	0.00%	87.50%	0.11%	1.87	1.00	No	No
volleyball	11.69%	65.92%	34.08%	14.21%	58.21%	71.98%	6.16	2.48	No	Yes
whatisthisthing	1.97%	64.46%	35.54%	1.93%	94.56%	5.89%	2.99	1.20	No	No

Quantitative Analysis of Reddit Communities

Table 2 presents the results from the Reddit crawls. Engaged users are defined as users creating a Reddit thread or responding to an existing thread within the timeframe examined. Exclusive Responders are users who respond to existing threads but never create threads. Initiators are users who create a new threads. Initiator-responders are users who have created at least one new thread and responded in a different thread which they did not create. Exclusive initiators are initiators who have never responded in a thread they did not create. Here, we only look at Focused and Critiques communities because subreddits were unlikely to differ across other tags due to the shared platform.

When the subreddits are sorted and split according to engagements per engaged user and initiators who also respond to other threads, 6/7 in the top-half are focused while only one in the bottom half is focused. When considering both engagements and engaged users per all registered users, the numbers are 5/7 and 2/7 for top and bottom halves respectively. When the subreddits are sorted and split according to

Table 3: Correlation analysis and significance tests between engagement stats and tag presence. Statistically significant relationships ($\alpha = 0.05$) are bolded.

	Engaged Users	EpRM	AE/EU	Initiators - Responder
Critiques	r = 0.143 p = 0.612	r = 0.100 p = 0.722	r = 0.180 p = 0.521	r = 0.284 p = 0.304
Focused	r = 0.503 p = 0.056	r = 0.479 p = 0.071	r = 0.538 **p = 0.039**	r = 0.707 **p = 0.003**

engagements per engaged user and imitators who also respond to other threads, 5/7 in the top-half feature critiques while only 2/7 in the lower half feature critiques. When considering engagements per all registered users the numbers are 4/7 and 3/8 for the top and bottom halves.

Based upon these findings, we conducted correlation analysis and significance testing, shown in Table 3. Focused subreddits showed more engagements per engaged user and more initiators responding in threads they did not create. These results suggest focused communities elicit more engagement than unfocused communities. No significant effects were observed for critiques.

CASE STUDIES

Third, we look closely at five communities: Fanfiction.net, StackOverflow, the learnprogramming subreddit, Wikihow, and DuoLingo. For each, we describe the community, list the tags from the tagging scheme, and describe the ways in which learning occurs at scale within the community.

Case Study: Fanfiction.net

Fanfiction.net is the largest online fanfiction site and community. Users with an account can write and publish their own fanfiction (original stories that use characters, settings, and other features from someone else's existing body of work) to the website. Other features include a built-in beta reader system, user-created forums, and user-curated archives of published works.

Tags: *Skill-building, Critiques, Content-Creation, Private Messaging, Forum-based Communication, Community Standards enforced by platform-moderators, Contains Sub-communities*

Part A: Learning by Creating

For each learning community, the first thing that needs to be identified is the content being learned. In the case of fanfiction.net, users who publish to the site are actively working to improve their writing and storytelling abilities by writing, posting, and soliciting feedback for their work. Communities such as these that revolve primarily around encourage users to develop a specific skill or ability are labeled **skill-building** communities in our tagging system.

Different skill-building communities may have different means of instruction or encouraging users to build their skills (for example, some may use tutorials to encourage users to practice and learn a specific skill). We consider fanfiction.net to be a **creation** based community, because users are specifically encouraged to improve their writing skills by, well, writing (i.e. creating content). Importantly, the platform depends entirely upon its users to create all the content (**artifacts**) for the site; without users creating and publishing their works the site would be content-less.

Part B: Learning by Consuming

All websites include content that is intended to be consumed. On fanfiction.net, content creators (writers) rely on others to consume (read) their work, however the relationship between creator and consumer within this community does not end there. The site highly encourages readers to leave **critiques** aimed at helping writers improved in the form of reviews posted a given work. While reviews are not strictly required to hold any specific educational value, a previous examination of reviews on Fanfiction.net did find that approximately 30% of the reviews in the sample provided positive feedback directed at specific elements of the text, 16.6% provided constructive criticism, and 5.4% provided a mixture of both (Evans et. al., 2017). Such critiques can provide an educational function similar to that of a peer review for both the reader and the writer, especially given that many readers are also writers on the site and vice versa. Because the act of consuming within this community can be tied closely to a specific learning purpose, we consider this to be a **content-consumption** based community, in addition to a content-creation based community.

Part C: Building a Community

On Fanfiction.net, writers (content-creators) and readers (content-consumers) are linked through the publishing-reading-review system as detailed above, but this link alone does not provide the foundation for the community. Rather, the platform provides additional features that allow users to connect with one another and create their own spaces for learning and socializing.

The platform extends on the review system through its beta-reader system, which establishes direct, longer-term relationships between individual writers and readers. Within this system, the beta reader agrees to act as an editor who provides more detailed feedback to the writer before the work is published to the site. The platform offers a portal for where readers can list themselves as "available" to beta read; writers

can then search the list for readers who might be a good fit, after which the two users can connect via **private messaging** to determine whether or not the relationship will work.

Users on the platform can always find like-minded users by reading and reviewing each other's stories or initiating conversation via private messages, but the site also allows users to create their own forums to make it easier to connect with others about a mutual interest.

Of these user-created forums, the single most popular is named "Writer's Anonymous" and was created for the express purpose of helping writers on the site improve via much more direct exercises. The forum is structured around a general critique thread and several technique specific "workshop" threads such as "Beginning/Opening Lines", "Death Scenes", "Dialogue", etc. where users can post excerpts pertaining to the technique to receive feedback from anyone. The forum has an additional Q&A component, since users can post their own new topics for more general questions about technique (rather than excerpts).

Explicitly the forum is meant to help writers improve their own skills, but implicitly the forum also teaches and encourages readers to become better reviewers. There are additional threads for "review games" designed to make sure that authors receive reviews that provide more meaningful feedback than "good job!" or "keep it up!" which are fairly common types of reviews to receive (Evans, et. al, 2017).

While this **sub-community** example is explicitly tied to wider community's primary learning target (writing skills), the user forum feature is very flexible topics are open-ended; many other forums are organized around socializing, interests unrelated to writing, or interests related to a specific fandom (i.e. fans for a given body of work). These sub-communities also have some autonomy enforce their own **community standards** (user to user etiquette) through **user moderators**. This differs from the main body of the site where abusive or otherwise inappropriate reviews or comments must be reported to and taken down by **platform moderators**.

Case Study: StackOverflow

Stack Overflow is a question and answer site for programmers. The users have built a crowdsourced library of detailed answers to many questions about programming. StackOverflow has a very strong gamification element, and it entices users to participate in the community by awarding points and badges and by having multiple leader boards. The points and badges earn "reputation" on the site, and that is used to unlock functionality. Reputation may only be various elements of the site, but as the site is well known by many software development professionals, having a high reputation on the site may also help a user in their career goals, as well as display topic mastery to the world.

There are many mechanics for earning reputation, but upvoting and downvoting are the primary mechanisms, where another user upvotes your contribution and you gain reputa-

tion, and if another user downvotes your contribution, you lose reputation.

Tags: Q&A main communication, Meta community present, Centralized community, Discussion-based learning, Private, Reward system present, Quality standards present, Community standards enforced by community, Critiques

Part A: Learning by Asking
Users who ask questions on StackOverflow are expecting other users of the site to help them. The site is built around a community that expects users to try to solve the problem themselves and then if they get stuck and need help, they can reach out by asking a Question. Questions are usually formed by posting the steps or work that you have completed and including what in your opinion went wrong, and finally asking the users of the site to guide you to a resolution. Even though users are building skills, this is not a site that was labeled **skill-building** by our tagging system, because the Questions are very broad and are not designed to build a single skill, but instead help you to solve problems and the learning is a consequence of that process.

Users generally learn by reading potential answers and comments posted to their questions. Comments often guide them to improve their asking of the question and that is often in the form of thinking about specific portions of the original question or general hints about asking questions on the site. Users learn how to use the site, but they are also learning about the topic they posted and the question they asked.

Part B: Learning by Answering
Users of the site can also learn by answering questions. The act of reading and thinking about other users' work helps that user to gain new knowledge and understanding and broaden their experience with the topic. Users typically monitor specific tags and as new questions are asked, the users will read the questions.

Interactions can then lead them to ask clarifying questions often in the form of a comment. That dialog between the asking user and the users of the community can lead to learning on both sides. They also interact with the asker by voting on the question. Upvoting a question is making the statement that this question has value to other users. Upvoting also increases the reputation of the asking user. Finally, they can interact by responding to the question with an answer. The act of teaching often improves learning, but in addition to that publicly answering a question opens the answerer and that response up to scrutiny.

Other users will examine the answer, often these users are users who feel they may have an answer to the question. So they can be harsh critics if the responses are not complete or incorrect. This critique also can lead to learning, again for all the community members involved.

Part C: Learning by Moderating
The final way to participate in the site does not involve asking or answering questions but is instead moderating the site.

Before you can moderate in any capacity, you need to gain enough reputation to unlock the basic moderator tools. These tools are part of the privileges that StackOverflow awards its users as they gain reputation.

By being able to moderate questions on the site, you begin to review questions not only for potential answers, but also you begin to enforce the community rules, and you learn and help other users to be better problem solvers by helping them to try before they ask, and then ask better questions once they have tried. You are also responsible for simply keeping the site valuable by moderating the questions and users of the site.

Finally, there are also moderators that are elected by users of the site. These moderators have the important job of following up on content flagged by the community. This helps to eliminate poor content and helps eliminate users of the site that are not improving the community.

Case Study: r/learnprogramming
r/learnprogramming is a subreddit forum focused on learning the nuances of programming, and covers a wide variety of topics across different languages and technologies. r/learnprogramming has the same overall posting mechanism and user features as the rest of Reddit, though it does have certain meta guidelines around how users are allowed to interact pertaining to keeping the community focused on learning the task at hand. Users can comment, and upvote/downvote different comments or entire post based on their perceived relevance to the r/learnprogramming community. Posts can be, Links, Questions, Advice, Videos, AMA, Freeform, or Actual code snippets.

Tags: &A main communication, Reward system present, Distributed knowledge, Direct (private) messaging, Forum based Communication, Community standards present, Tag system present, Community standards enforced by community, Community standards enforced by software/platform, Quality standards present, Quality standards enforced by system, Discussion-based learning, Focused, Critiques

Part A: Learning by Asking and Creating
Users of r/learnprogramming can do a variety of things, largely falling under asking specific questions or creating content like guides and videos. The community is focused on learning various topics related to programming, and in a similar manner to StackOverflow, expects well thought out questions that are the result of self-exploration.

Questions have a specific structure, and can fall into two different broad categories: conceptual and debugging related questions. Debugging related questions are largely treated in a similar way to the questions asked on Stack Overflow. Users are expected to form debugging questions as a result of some specific process or problem, and are expected to provide context around the question.

Because the subreddit's overall goal is to facilitate the learning of programming, debugging is a topic that tends to arise

fairly often. Contrary to other communities, users are actively encouraged to ask deterministic debugging questions for even the smallest minutiae. There are a handful of guidelines that are enforced in the way these posts are structured. Users are expected to provide all contextual information in the form of a good description and include a descriptive, short title. A minimal example program exemplifying the problem should also be included, along with error messages.

There is a plethora of documentation around posing and answering questions the correct way included in the subreddit as well. Conceptual questions are asked in a similar frequency to specific debugging questions. Community members are requested to check out older posts first to maintain a sort of master set of questions. If a new question is posed based on an older one, it must specify why the older one did not satisfy that question. Esoteric questions and niche topics are also allowed here, though the community guidelines explain that the members may not have sufficient expertise to answer these. Users can also create content in the form of guides or videos that explain a specific topic, or include learning programs and curricula.

Learning is done when users read responses, answers, and comments to their questions and other posts. These comments guide them to prune and improve their questions, point out learning opportunities, and focus on breaking down reasoning into digestible chunks. Discussion is encouraged and fostered through the site features.

Part B: Learning by Answering and Critiquing.
r/learnprogramming users also learn by answering questions in a similar manner to other Q&A communities. Reading and discussing other user problems, work, or content helps users solidify their own knowledge or get introduced to new topics. This is especially true in the conceptual questions, as they are often open ended and tend to start longer form discussions. Users will often track posts that have interesting keywords or tags, and come back frequently to see new comments posted. Through interaction with other users, a specific user can learn in the form of explanation or asking for clarification. Users are also able to vote for better posts, answers, and comments, leading to curation and meta learning in the community. In this way, there is an element of teaching in the learning process on r/learnprogramming as well.

Part C: Learning by Moderating and Building a Community
Finally, users can also learn and participate in the community through moderating posts and observing which posts improve with time and which do not. Moderating requires a reputation, measured above a certain karma setting on the Reddit platform. Moderation allows individual learners to contribute to building the community, and also learn which posts make for useful learning content. Moderators can also be selected by the community in a similar way to other subreddits and question answer sites.

Case Study: Wikihow
Wikihow is an online learning community built around the creation and distribution of how to guides. Topics are extremely diverse, posts are broken down into a step by step format, often along with image to guide learners. Effectively, the knowledge base is a crowdsourced database of guides that are easily searchable through the site interface. Posts can be curated through a rating feature, building a prioritized feed across the site. Wikihow has some interesting features, including some gamification of user profiles and comment threads. Other features include user bios, article creation, editing, discussions, profile discussions, review of proposed edits, positive reviews can be used as an article banner to demonstrate credibility, anonymous community question and answer section per article, flagging features, and non exclusive user groups.

Tags: *Teacher Present, Skill building, Critiques, Artifacts, Focused, Distributed knowledge, Community standards present, Community standards enforced by community, Community standards enforced by software/platform, Reward system present, Quality standards present, Quality standards enforced by system, Tag system present, Dispersed community, Creation-Based Learning, Tutorial-based learning*

Part A: Learning through Reading
The primary function of Wikihow is to present users with a plethora of community created how to guides. Correspondingly, much of the learning on the platform is done through users reading guides, trying them out, failing or succeeding, and repeating the cycle. Guides facilitate ease of comprehension through a broken down step by step list framework that has parallel visual examples as well.

Part B: Learning through Writing
Users that contribute articles to Wikihow learn from the process as well. This learning can be directly about the topic by learning a topic more deeply as they consider how to explain the topic to other users, and also through feedback about the article's content and quality from the community. Contributors on the platform also learn about the act of writing useful guides, and through the feedback of other community members. Various rating, flagging, and editing review features allow for this feedback to be incremental in the writing process, and allows for learning with each writing iteration. In this way, the collaborative work features also teach newer members of the community what constitutes good writing and useful content.

Part C: Learning through Questions
Members of the Wikihow community can also learn in another directed fashion: question and answer threads. Articles often have specific question and answer threads that allow users to anonymously ask questions about the content of the article, allowing them to come back with feedback or questions about areas in which they did not achieve the intended result of the guide. This feedback loop leads to a continual

learning process on the part of both the contributor of the article and the user reading the articles.

Part D: Learning through Review and Moderation
Moderation, of both article content and quality, is another way to learn from the Wikihow site. Moderation exists at various levels of the site's functionality. This includes reviewing specific edits made by contributors to articles, flagging articles for violating community guidelines, rating articles based on their content, and correcting answers in the question and answer sections. Moderation makes up a huge part of the community and drives a large part of the site's curation and involves all users.

Case Study: DuoLingo
DuoLingo is a language learning platform representing and blending the features of a MOOC and a gamified informal learning community through its forum platform. DuoLingo's forum revolves around users sharing, seeking, and producing content and knowledge relating to the primary DuoLingo teaching content and often related lessons as well. The site has standard forum features, with post creation options and reply functions where other users can interact with the original poster and other members of the community. Importantly, the community is categorized by languages, so topics do not have to mix and users trying to learn a specific language can fit into a sort of subcommunity within the forum. In this community, users can start discussion threads, respond to other threads, curate a given thread or response, upload an avatar to their profile, and include a simple tag line or header in their profile.

Tags: *Skill building, Forum based Communication, Direct (private) messaging, Community standards present, Community standards enforced by community, Community standards enforced by software/platform, Reward system present, Quality standards present, Quality standards enforced by system, Tag system present, Discussion-based learning, Tutorial-based learning*

Part A: Learning through Posting
Users that contribute posts to the forums on DuoLingo can learn through the process in a variety of ways. Many of these posts are questions about the topic language's grammatical form or speech patterns, and in asking these questions, users usually have to discretize and scope out what they are trying to learn, leading to better understanding. Along with asking questions, users can create a post about anything else, including creating pieces of useful content like tip guides, recommending related resources or sharing videos about the topic. Users learn through pruning their posts and getting feedback from the community around the content or questions they share.

Part B: Learning through Discussion
Because it is primarily a forum, users can learn on the Duolingo forum platform through replying and discussing the topics that are shared as well, often exchanging different examples and experiences around how to pronounce words, what the correct forms of certain words are, how different

subcultures and dialects speak, or write, and general habits of associated cultures.

CONCLUSION & FUTURE WORK
In this paper, we have presented an effort to make sense of informal learning communities at scale. While the learning at scale literature has largely focused on MOOCs, online degree programs, and tools to support more formal educational experiences, massive informal learning communities have gone relatively unstudied in this context. In order to interpret these communities at scale, we have devised a tagging system based on features related to the platform and social structures present within the community. Further development in this area could include correlation studies to determine if any of these tagged features tend to cluster together, quantitative studies to further relate engagement with different tagged features, and qualitative studies to gauge user perception of the various tagged features. There is also room to apply this research toward specific subsets of online learning communities, including more formalized learning environments and distance learning, and communities of practice. Because learner characteristics were excluded from this review, further research should be conducted to gain insight on learner demographics, expertise, motivation, etc. and to understand the relationship that exists between learner and community characteristics.

Quantitative analysis beyond the subreddit communities should be conducted to include all the publicly accessible communities tagged. This wider analysis could identify potential effects of additional community and platform characteristics. Data from a larger time frame than a year could be considered. This longer timeframe could provide more detail about community engagement and the life-cycle of communities. If characteristics eliciting greater community involvement are identified, informal learning communities could be optimized to elicit greater engagement.

Additionally, qualitative studies could involve coding user interactions relative to any of the tagged features in order to better understand the ways in which these elements can shape community culture. Interviews could be conducted with participants in communities sharing the same tagged feature to see if any elements shape user perceptions of learning, feelings of belongingness, and any other metrics regarding the effectiveness of a learning community to further shape community design for future platforms.

In short, there is ample future work that can and should be done in understanding informal learning communities at scale; the domain is more difficult to analyze than formal course experiences, but the learning observed may go beyond those carefully-curated offerings. Arguably the most remarkable instance of learning at scale is the fact that a small number of dedicated people along with open-source infrastructure can construct, without video lectures or automated evaluation or AI teaching assistants, learning experiences that touch millions of lives.

REFERENCES

[1] Aoki, K., Fasse, R. & Stowe, S. (1998). A Typology for Distance Education-Tool for Strategic Planning. In *World Conference on Educational Multimedia and Hypermedia & World Conference on Educational Telecommunications.*

[2] Aragon, C. & Davis, K. (2019). *Writers in the Secret Garden: Fanfiction, Youth, and New Forms of Mentoring.* MIT Press.

[3] Chuang, I. & Ho, A. (2016). HarvardX and MITx: Four years of open online courses--fall 2012-summer 2016. *Available at SSRN 2889436.*

[4] Colley, H., Hodkinson, P. & Malcolm, J., (2002). Non-formal learning: Mapping the conceptual terrain. A Consultation Report [White Paper]. Retrieved from http://infed.org/mobi/non-formal-learning-mapping-the-conceptual-terrain-a-consultation-report/

[5] Czerkawski, B. C. (2016). Blending formal and informal learning networks for online learning. *International Review of Research in Open and Distributed Learning, 17*(3), 138-156.

[6] Eastmond, D. V. (1998). Adult learners and Internet-based distance education. *New Directions for Adult and Continuing Education, 78*, 33-41.

[7] Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology, 16*(2), 171-190.

[8] Folkestad, G. (2006). Formal and informal learning situations or practices vs formal and informal ways of learning. *British Journal of Music Education, 23*(2), 135-145.

[9] Hrastinski, S. (2008). Asynchronous and synchronous e-learning. *Educause Quarterly, 31*(4), 51-55.

[10] Johanes, P. (2019). Start of a Science: An Epistemological Analysis of Learning at Scale. In *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale.* ACM.

[11] Lowenthal, P. R., Wilson, B. & Parrish, P. (2009). Context matters: A description and typology of the online learning landscape. Paper presented at the 2009 AECT International Convention, Louisville, KY.

[12] Malcolm, J., Hodkinson, P. & Colley, H. (2003). The interrelationships between informal and formal learning. *Journal of Workplace Learning, 15*(7-8), 313-318.

[13] Marsick, V. J. & Watkins, K. E. (2001). Informal and incidental learning. *New Directions for Adult and Continuing Education, 2001*(89), 25-34.

[14] Morgado, E. M., Yonezawa, W. & Reinhard, N. (2003). Exploring Distance Learning Environments: A Proposal for Model Categorization. In *International Academy for Information Management (IAIM) Annual Conference: International Conference on Informatics Education & Research (ICIER).*

[15] NSTA (National Science Teachers Association). (1999). Position statement on informal science education. Retrieved from http://www.nsta.org/about/positions/informal.aspx

[16] Porter, C. E. (2004). A typology of virtual communities: A multi-disciplinary foundation for future research. *Journal of Computer-Mediated Communication, 10*(1).

[17] Riel, M. and Polin, L. (2004). Online Learning Communities: Common Ground and Critical Differences in Designing Technical Environments. In Barab, S., Kling, R., and Gray, J (Eds.) *Designing for Virtual Communities in the Service of Learning.* Cambridge University Press.

[18] Roberts, T. S., Romm, C. T. & Jones, D. (2000). Online Courses and Collaborative Learning: Underlying Philosophies and Practices. In *International Academy for Information Management Annual Conference.*

[19] Stanoevska-Slabeva, K. & Schmid, B. F. (2001, January). A typology of online communities and community supporting platforms. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences.* IEEE.

[20] Vrasidas, C. (2002). A working typology of intentions driving face-to-face and online interaction in a graduate teacher education course. *Journal of Technology and Teacher Education, 10*(2), 273-296.