QMaps: Engaging Students in Voluntary Question Generation and Linking

Iman Yeckehzaare
School of Information
University of Michigan
Ann Arbor, MI, USA
oneweb@umich.edu

Tirdad Barghi
Independent Scholar
Tehran, Iran
tirdad.barghi@gmail.com

Paul Resnick
School of Information
University of Michigan
Ann Arbor, MI, USA
presnick@umich.edu

ABSTRACT
Generating multiple-choice questions is known to improve students' critical thinking and deep learning. Visualizing relationships between concepts enhances meaningful learning, students' ability to relate new concepts to previously learned concepts. We designed and deployed a collaborative learning process through which students generate multiple-choice questions and represent the prerequisite knowledge structure between questions as visual links in a shared map, using a variation of Concept Maps that we call "QMap." We conducted a four-month study with 19 undergraduate students. Students sustained voluntary contributions, creating 992 good questions, and drawing $1,255$ meaningful links between the questions. Through analyzing self-reports, observations, and usage data, we report on the technical and social design features that led students to sustain their motivation.

Author Keywords
Question Generation; Concept mapping; Learnersourcing; CSCL; Collaborative Learning; Intrinsic Motivation; Learner-Centered Design

CCS Concepts
•**Human-centered computing** → *Social content sharing;* **Collaborative content creation; Computer supported cooperative work;** *Collaborative interaction; User centered design; Participatory design;* •**Applied computing** → **Interactive learning environments; Collaborative learning;** *Distance learning; E-learning;* Computer-assisted instruction;

INTRODUCTION
Many studies report that multiple-choice question generation by students positively impacts deep learning [64, 19, 76, 25, 16, 65, 14, 6, 27, 47, 45, 4, 8, 17, 60, 49, 22]. Visualizing relationships between concepts also helps students to understand the relations among concepts and identify new relations, what Ausubel calls meaningful learning [2, 51, 53]. However, these learning activities are challenging for students, and it is hard to maintain their motivation to do the activities. Denny et al. [15] reported on the voluntary use of a question generation system, PeerWise, in a semester-long class; less than 5% of students created any questions. Similarly, we are not aware of any study of concept mapping where students showed motivation to generate links between concepts over an extended period of time without a graded assignment that required it.

Our main contribution in this paper is to identify a set of design features that motivate students to engage in these difficult but valuable learning activities voluntarily. Over four months, 19 undergraduate students engaged in a not-for-credit[1] learning process to study Python and HTML/CSS. Collectively, they generated $1,154$ questions of which the instructor deemed 992 to be of high quality, and drew $1,255$ high-quality relations between questions. Some of the innovative design features that seemed to contribute to sustaining motivation to participate in these challenging learning activities voluntarily are:

- **QMap of relations between questions**: when adding any new question, students had to identify at least one existing question as a prerequisite of the new question. Relations were added to a variant of concept maps [53] that we designed and called *QMap*.
- **Identifying question authors**: peers could see which student had created each question.
- **Ownership transfer with justified edits**: a student could "claim ownership" of an existing question by modifying it and providing a text justification for the modification.
- **Visible log of activities**: contributions can get lost when they are dispersed throughout the QMap. Students saw a linear log of all activities in a Slack channel.
- **Continuous assessment**: both peers and the instructor commented and voted on the quality of questions and relations.
- **Instructor role modeling**: the instructor also generated some questions and relations.

BACKGROUND
Crowdsourcing in learning science has been realized in the form of collaborative content generation by students, usually called "learnersourcing." Successful examples include

[1] One student elected to earn one credit-hour of independent study.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376882

crowdsourcing interactive concept mapping of steps in how-to videos [43], labeling sub-goals in videos [72], generating explanations [74], personalized hints [24], step-by-step video annotations [36], video traces [35]. Learnersourcing is also used in the generation and evaluation of questions and solutions by students for exams and practice [71]. Our focus in this study is on learnersourcing multiple-choice question generation, evaluation, mapping, and continuous assessment. We begin with prior research on learning theories and activities that informed some of our design goals and design features.

Collaborative Question Generation

Collaborative question generation improves self-confidence and critical thinking, sheds light on different ways to solve a problem, supports ideation, exhilarates students, facilitates rapid peer-evaluation, and improves deep learning [64, 19, 76, 25, 16, 65, 14, 6, 27, 47, 45, 4, 8, 17, 60, 49, 22]. Creating question stems, correct/wrong choices, and justifications for each choice involves critical thinking and evaluation that result in deep learning [25, 49, 20]. Multiple-choice question generation also requires metacognitive thinking about why other students might choose any of the correct or wrong answers [20]. When designing convincing wrong choices, students pay attention to misconceptions and innovate their own solutions to those mistakes, which enhances their deep learning [25].

PeerWise [19, 13] is a popular online system for collaborative multiple-choice question-generation. It provides students with a learning environment where they generate questions and share them with their classmates for self-testing. Question-generation through PeerWise is found helpful in a large spectrum of disciplines [64, 25]. Replacing weekly course assignments with question generation also showed a significant increase in final exam grades [4]. Moreover, question generation is shown to improve not only exam performance but also students' engagement in more learning activities [25, 6].

Evaluation of Collaboratively Generated Questions

Sustainable assessment should be integrated into lifelong, lifewide, and life-deep learning, such that learners get involved in designing, scheduling, evaluating, and reflecting on their own assessment [9]. Peer-evaluation is shown to facilitate and enrich formative assessment. Piaget identifies disagreements between peers as a stimulus for metacognitive reasoning to find convincing resolutions [48]. Peer-evaluation of collaboratively generated questions is helpful in that peers are uncertain about the correctness of the questions designed by their peers and engage in metacognitive thinking to resolve the uncertainties. So, it is necessary to provide students with constructive feedback on the generated questions and encourage them to discuss the questions with their peers [17]. Commenting on generated questions in PeerWise by students was correlated with increased exam grades [19]. While assessment of questions in PeerWise is assigned merely to peers, it is advised to conduct peer-evaluation under the supervision of an instructor, especially in the early stages, to sustain students' motivation [29] and prohibit possible collusion [68].

Motivation in Collaborative Learning

Students' motivation to create questions significantly affects their learning [42]. This motivation typically declines over time [33, 31]. Different studies have incentivized students' contributions through course grades or paying them as subjects of the study [76, 14, 25, 29, 20, 49, 38]. Gamification, through a serious board game, is also proposed [77]. Contrarily, Denny et al. [15] randomized students in a physiology class into four experimental groups with different incentives: "control," "points," "badges," or "both," to generate and practice multiple-choice questions. While badging significantly increased self-testing (question answering), especially among high-achievers, none of these experimental conditions showed any significant effect on question-generation activities.

Self-determination theory (SDT) classifies the source of motivation into intrinsic (e.g., enthusiasm or joy of learning) and extrinsic (e.g., grades, points, or badges) [58, 59]. Research shows that students with higher intrinsic motivation exert higher effort and achieve better learning, retention, and transfer [57, 58, 59]. SDT further categorizes extrinsic motivation along a continuum from "external regulation" (most controlled) to "integrated regulation" (highest autonomy). It identifies three factors that contribute to intrinsic motivation:

- **Competence**: one's belief in the ability to perform a task,
- **Autonomy**: ability to decide how/when to perform a task,
- **Relatedness**: belief about how others value a fulfilled task.

SDT identifies autonomy as the most essential factor and urges the provision of autonomy rather than external rewards [59]. It explains that while positive feedback and encouragement improve students' perceived competence in learning, external rewards may stimulate a sense of loss aversion or being controlled, which due to the "overjustification effect," may crowd out intrinsic motivation [41, 59].

Too much autonomy, however, may lead students not to participate very much, for two reasons:

- **Lack of self-regulation**. Many students lack regulatory skills [55]. Self-regulation is needed to persist when learning activities are challenging [62].
- **Social loafing** is defined as free-riding that leads to under-provision of public goods [54, 79, 69, 61, 32, 40]. Denny et al. [15] report that 701 students participated in their study and answered $50,759$ questions, but only generated 376 questions. On average, each student answered 72.4 but generated 0.54 questions, which demonstrates social loafing. The Collective Effort Model (CEM) [32] suggests several factors that may reduce social loafing when collaboratively producing public goods. These include allotting intrinsically meaningful tasks, informing individuals of the importance of the task, providing feedback about their own or group performance, and making salient the impact of each participant's work on the collective outcome.

Concept mapping and meaningful learning

In the question generation process, students may solely focus on designing questions and neglect investigating conceptual relationships between new concepts and what they previously learned. Assimilation theory [52, 53, 3, 2] states that "learning takes place by the assimilation of new concepts and propositions into existing concept and propositional frameworks held by the learner" [53]. "Meaningful learning" happens when "learners "integrate" new information into old information" [2]. "[c]reativity results from very high levels of meaningful learning" [53]. Rote learning happens when students only focus on the concepts they are learning at the moment; in contrast, meaningful learning occurs when they relate the new concepts to those they learned before [53, 52, 51, 37].

One way to encourage students to assimilate new concepts is to have them place the concepts in a visual map. We designed *QMaps*, a variant of Novakian concept maps. We provide an overview of the concept mapping literature in this subsection; QMaps are described later.

"Concept mapping" has three building blocks: concepts (nodes), relations (linkages), and propositions [53, 12]. In Novakian concept maps [11], relations have linking words (short labels), nodes consist of few words, maps are small, and "string maps" are not favored. They have a "hierarchical organization, with the most general, most inclusive concepts at the top, and progressively more specific, less inclusive concepts at lower levels" [12, p. 8]. Students need "expert skeletons," small concept maps prepared by experts, as scaffolds to further knowledge construction and expand the maps [52]. In contrast to concept maps, the linear structure of textbooks may result in learning independent chains of concepts, which diminishes meaningful learning [37]. The relations drawn between concepts by students were originally used to *measure* their meaningful learning [12, 11, 37], but engaging in concept mapping is also thought to *cause* meaningful learning.

Concept mapping's inherent focus on meaningful learning results in higher test scores [52]. In a longitudinal study conducted in a high school in Costa Rica, the average approval rate in the national high school exam was 56% to 81% from 2000 to 2002. By adapting concept mapping in their curriculum, after five years, 100% of students passed the exam [52].

Social constructivism explains that language and social exchange can enhance cognition development [73]. Collaborative concept mapping, as a social exchange, can aid in correcting misconceptions because peers are at similar Piagetian developmental levels and better understand each other [50]. ConceptScape is a collaborative system to facilitate interactive concept mapping of videos to improve understanding and encourages the participation of novice students to generate high-quality concepts collaboratively [43]. However, when collaborating remotely, understanding others' changes to a concept map is distracting and difficult. Collaborative concept mapping has been shown to be beneficial only when performed synchronously [5]. Some studies also report that students prefer synchronous over asynchronous concept mapping [10, 63].

METHODOLOGY

We iteratively designed tools and processes for students to learn through creating and mapping questions collaboratively. Over sixteen weeks of summer 2018, nineteen undergraduate

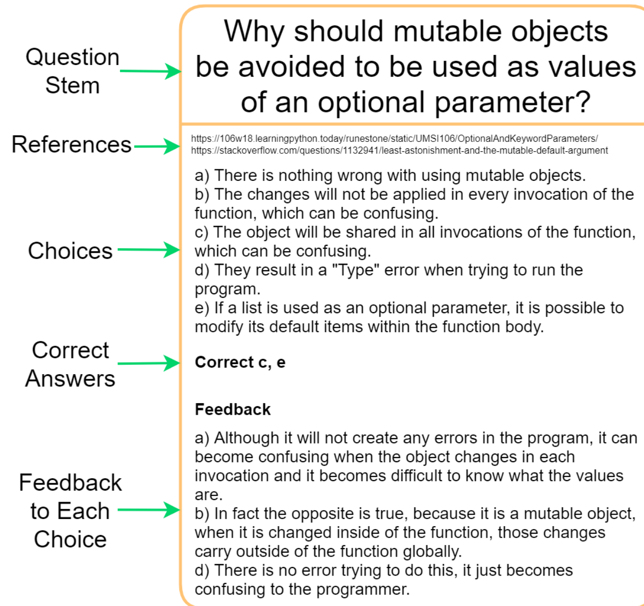

Figure 1. A sample node representing a multiple-choice question.

students at the University of Michigan, School of Information, including fifteen women and four men (19-22 years old), participated in our study. Students were invited to learn about python and HTML/CSS. We only invited students who had previously received *A* grades in at least one programming course; most had taken an introductory programming course in python that was intended for non-majors in computer science. Following SDT, we thought that students who felt competent after earning a high grade in a course would have the most motivation to continue learning in a self-directed way. Two authors of this paper participated as instructors. One of them had served as a teaching assistant leading a section in the introductory python programming course. One of the nineteen students was from that section and thus knew the instructor well.

Students and instructors interacted online only; the first in-person meeting was a retrospective group reflection session at the end of the summer. For all the online interactions, we used free versions of "Draw.io" for collaborative question generation and visualization, and "Slack" for group communication and voting. Students self-explored various online resources to learn and generate questions to share on the Draw.io map. Each question consisted of a stem, two or more correct/wrong choices, and feedback explaining each choice (Figure 1). Students drew arrows on the map in Draw.io to indicate prerequisite relations (Figure 2). Whenever a student added or modified a question on the map, they shared it on the Slack channel to be evaluated by others (Figure 3). Through the summer, we created and refined conventions for using these off-the-shelf tools. We refer to these conventions as design features and will discuss them in the next section.

The instructors played an essential role in supervising and evaluating students' learning activities, and teaching them how to explore, filter, and learn through the abundant online resources that were available to them. They voluntarily extracted knowledge from multiple resources, filtered them, and shared the essence of what they learned with their peers through creating multiple-choice questions and relating them to each other. They even found helpful websites that we were not aware of. We believe this process helped the students to learn how to use Internet resources efficiently for learning.

DESIGN FEATURES

Our design features are inspired by Self-Determination Theory and the Collective Effort Model. SDT asserts that students' motivation comes from autonomy, competence, and relatedness. CEM asserts that identifiability of work and increasing the salience of individual and group benefits will increase motivation to contribute to public goods.

In this section, we describe features of the learning process, with a particular focus on how they contribute to these goals. Autonomy is a core element of our design: students choose what to study, and when, where, and how much to participate. Other goals are served by particular design features discussed below. For some of the design features, we describe iterative changes that were made during the first month of the study, in response to students' feedback and our observations.

Organized QMap of Questions

We provided students with a concept mapping tool to collaborate on visually relating prerequisite questions:

- Each node represents a complete question, including a stem, available choices, and feedback for each choice (Figure 1).
- To facilitate QMap navigation, nodes can be in two modes: **open** (all question parts are visible) or **closed** (only question stem is visible) (Figure 2).
- To reduce reader disorientation and improve readability [39, 44], QMaps can be zoomed in/out, and the node titles (question stems) and linking words on relations are in a consistent font-size, larger than the other node parts (Figure 2).
- When adding a new question node, a student must draw one or more links from prerequisite questions, and label the links with words or short phrases.
- As QMaps grow in size, understanding the relations between questions can become confusing, especially for newcomers. To mitigate this problem, we constrained all the nodes to have the same width and aligned them vertically in columns with equal spacing between columns. In our study, we maintained these formatting constraints through a Wizard-of-Oz approach: the instructor manually edited whenever students did not follow the convention.

As an aside, readers who are very familiar with Novakian concept maps may want to note a few contrasts with our Qmaps. QMaps progress to more complex material from left to right, while Novakian concept maps read top to bottom. Nodes in QMaps contain complete questions, while in Novakian concept maps, nodes are supposed to have no more than three words. Relations in QMaps can be labeled with any phrase, while in the Novakian concept maps, node words and link words should be composable into propositions.

Placing and linking new questions on the map is a learning opportunity. It forces a student to:

- connect a new question with existing knowledge, promoting meaningful learning.
- develop a mental model of the topics in different regions of the overall map, which also promotes meaningful learning. Without such a model, students would have to engage in exhaustive search through all the existing questions in order to place a new question.
- fully understand a question if they found it in an external source, in order to correctly place it on the map.
- study peers' questions, promoting collaborative learning.
- notice if the question they are adding is redundant. In unordered collections of content, it is hard to know whether new content is redundant. In a map, once one finds the right place to put a potential new question, it becomes apparent if it duplicates an existing one.

In our study, students collaboratively created two QMaps, one about Python and another one about HTML/CSS. A portion of the latter is shown in Figure 2. Initially, students had difficulty with introducing new topics on the map. This observation reminded us of the necessity of expert skeletons in concept mapping. So, we changed the design such that the instructor initiated new topic questions and relations as scaffolds, allowing students to expand by adding offspring questions about use-cases and more detailed/advanced topics.

Feedback and Voting

To discourage procrastination and social loafing, our design allowed students to write comments and vote on the quality/helpfulness of others' contributions. All the feedback and communication were conveyed through Slack channels. After adding a question to the QMap, a student posted the question stem into Slack. Others then used emojis to upvote and started conversation threads. If a student found a question while browsing the QMap long after the question was first posted, they were taught to search the corresponding Slack channel for the stem and vote or comment on it, which was a somewhat cumbersome process.

Ownership of Questions is Identified

Most studies on question generation by students have proposed the anonymity of contributions. Conversely, the Collective Effort Model advises making contributions "identifiable" to reduce social loafing. Similarly, in computer-mediated communication tools, students usually dislike Public Anonymity and Private Accountability (PAPA) [18]. Thus, one of our distinguishing design features is the lack of anonymity. Since Draw.io does not afford the identification of question creators, we asked students to copy their question stems to Slack, which did display their names (Figure 3).

Ownership Transfer with Justified Edits

The design features mentioned above incentivize students to create high-quality questions. However, research has shown that members of online communities may hesitate to edit each

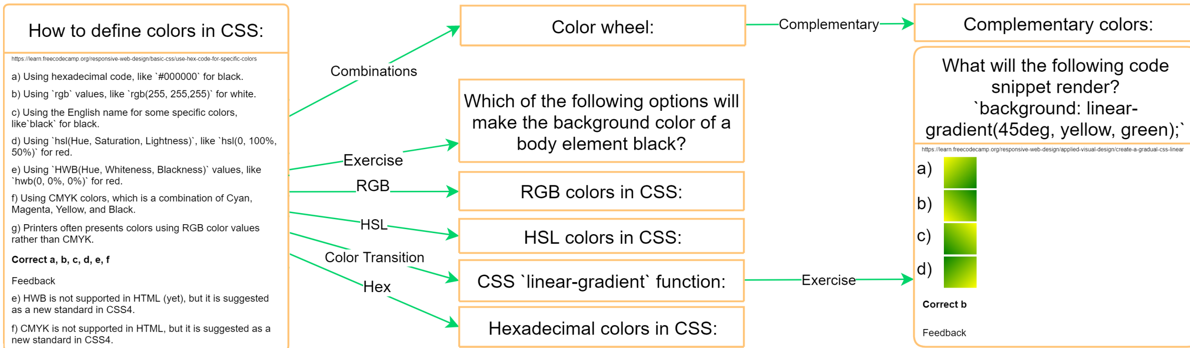

Figure 2. A small section of the QMap with CSS content. There are 7 closed and 2 open (expanded) nodes, and 8 relations.

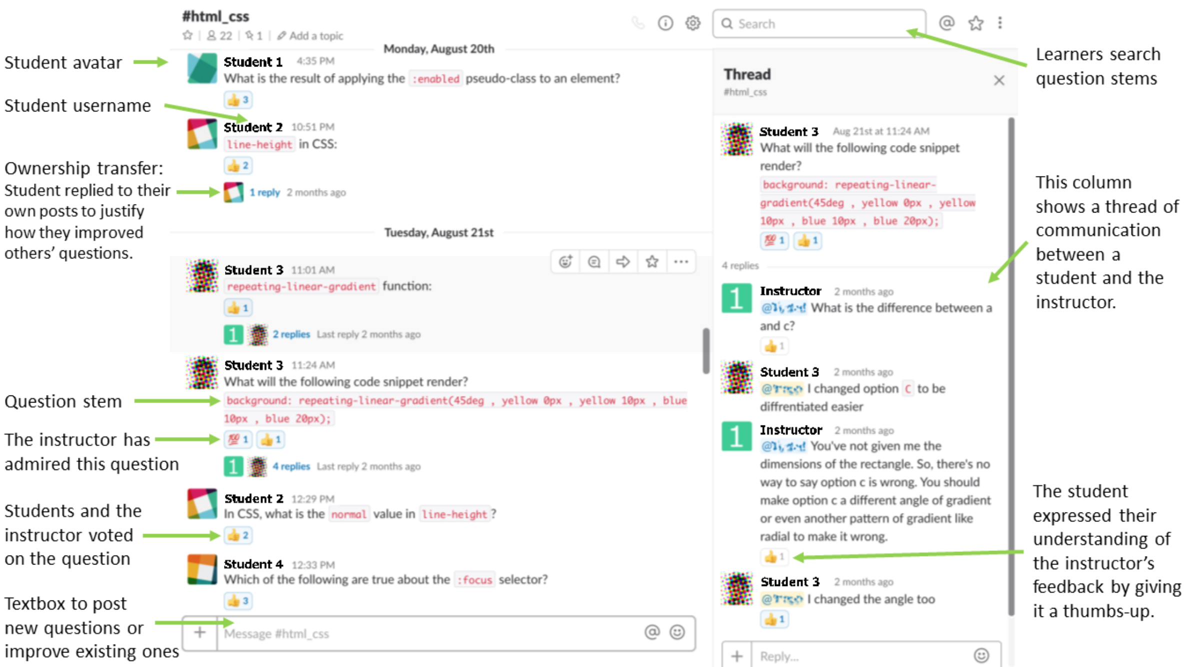

Figure 3. Screenshot of the Slack channel showing multiple-choice question stems, communication, constructive feedback, and ownership transfer. The students and instructor names are replaced with "Student #" and "Instructor" for privacy reasons.

other's contributions [26]. To solve this issue, a student who edits a question or its relations with other questions on the QMap posts their changes to Slack. Thereafter, any student who discovers the question in the QMap and searches for it in Slack can find both the original post and the latest post. Thus, both parties would be perceived as having contributed to the question. Among students and instructors, editing a question and making a post about it was referred to as "claiming a question," and the last person who had claimed the question was referred to as the "owner." This feature incentivized not only improving questions but also creating high-quality ones to ensure their ownership would not be transferred to others.

As a requirement for claiming the ownership of student *A*'s question, student *B* was supposed to post justifications for their modifications to the Slack channel. We thought that would help *A* understand their possible misconceptions or mistakes, and why and how their question was modified. The instructor and peers verified the justifications, and if any of them was not convincing enough, the instructor reverted the question to its previous version and owner.

Privately Revealing Cumulative Scores

Self-determination theory accentuates feeling competent as a way to sustain students' motivation. Seeing vote counts on each posted question may give the student a sense of competence, which can be augmented by a reputation system that computes and displays an aggregate measure of performance and achievements [56]. In our study, we used the Slack API to compute the total votes each student received. However, we revealed the aggregate reputations privately to each individual only once, in a personalized email on 07/16/2018, as a "nudge" to motivate contribution without making point accumulation a focal point throughout the process. Public revelation of reputations would have allowed for social comparisons. However, people mostly compare upward, and the common response is contrast. Upward comparison decreases self-evaluation, which is more severe in highly selective institutions [23]. While each member could roughly estimate these cumulative numbers based on the publicly visible votes on each question, by not revealing aggregate reputations publicly, we made such comparisons more difficult.

Visible Log of Activities

Sun et al. [67] showed that observing peers' current contributions and interactions are essential for sustaining students' motivation and satisfaction. Following "matching of effort" [28] rather than "social compensation" [75] theories, we thought that visible signs of life from other students were more likely to help maintain students' motivation than they were to lead to social loafing. However, when people collaborate on a single collective document where edits can be made in many places (in our case, the QMap in Figure 2), it can be difficult to follow others' actions, or indeed even to know whether there has been any activity. For example, it can be challenging to know what has changed on a Wiki page or a Google Doc [70]. A similar problem has been noted in discussion forums with deep threading, where new contributions may be difficult to notice. The usual solution is some kind of chronological log of activity [34, 70]. In our case, the Slack channel where students posted copies of question stems served as that log (see Figure 3).

Topic Sprints

When students can write questions about any topic they like, they may be inclined to design questions only from the topics they have understood well [16]. To encourage broader exploration, Denny et al. [16] suggested assigning randomized topics to students to generate questions from. However, this may cause another issue, that students will focus on one assigned topic and miss other topics.

In response to student requests early in the semester, we designated a new topic periodically with a suggestion that everyone focus on it together. Sometimes the topic was defined by a specific external resource. For example, on July 23, a group email went out that included the following: "For this week, I suggest finishing https://learn.freecodecamp.org/responsive-web-design/basic-html-and-html5/". Topic sprints have also been shown to be helpful in past work [16]. To maintain autonomy, however, we allowed students to work on any topic at any time. Some students chose not to join the group on the current topic. Indeed, one student kept working on Python for the whole summer, even after most of the other students moved on to HTML and CSS topics.

Continuous Instructor Assessment

Instructors assessed the quality, correctness, and helpfulness of all parts of each question (stem, multiple choices, and choice-specific feedbacks) and relations between them. Students could revise their generated questions, votes, and comments multiple times based on the feedback they received (see left column of Figure 3). In contrast to periodic exams, this enabled continuous assessment of each student's learning and development through the study. An instructor's supervision is crucial to identify and correct students' misconceptions, which is consistent with prior studies on improving peer-evaluation [29, 68, 46]. Similarly, receiving personalized feedback from the instructor about the relations drawn in a concept map is shown to improve learning [43]. Also, continuous assessment improves students' engagement in social learning and may result in higher levels of learning in Bloom's taxonomy [25].

Role Modeling

The instructors contributed to the community in the same way students did to model appropriate norms. Previously, the pivotal impact of role modeling on residents' learning and motivation was studied [7, 66]. Role modeling is referred to as a "hidden curriculum" that helps students learn by observing high-performing peers and receiving feedback from them [30]. Also, mentors' engagement in online communities can improve the quality of novice members' contributions [21].

EVALUATION OF OVERALL OUTCOMES

First, we assess whether the process and tools as a whole succeeded at maintaining student participation in the challenging learning activities of generating and relating questions. As described in the background section, prior research theorizes that these activities lead to positive learning outcomes, so long as the students generate high-quality questions and relations. We do not attempt to validate the theories by separately measuring learning through an exam or other means. Instead, our primary outcome is the quantity and quality of the questions and relations students generate, and whether the quantity is sustained over time.

Question Quantity

Over the four months of the study, 19 students voluntarily created $1,154$ questions. The average number of questions per week declined gradually, but still averaged more than three questions per student even in the second to last week (Figure 4). We also note that it took more time to create questions later in the study: when a map got larger, it was more challenging to create questions that were conceptually distinct from the existing ones and to identify prerequisite questions.

Question Quality

The questions were generally of high quality. Of the $1,154$ questions that students created, 992 (86%) were above the minimum quality threshold to earn an instructor upvote, indicating that the instructor thought they would be appropriate to administer to students in a regular course as learning or assessment questions. Among them, 221 earned an "instructor admires" badge, indicating exceptional creativity or making subtle connections to other concepts. Average quality improved over the 16 weeks as students learned from the instructor and peer feedback what makes a good question (Figure 4 (b) and (c)).

Prerequisite Relations Quantity

The students identified $1,255$ prerequisite relations between questions, an average of 1.09 per question. Relating new questions to existing questions did not come naturally for all students. Early in the semester, the instructor often gave constructive feedback, requesting that students add or revise the relations they had drawn to existing questions. After the first two weeks, however, the instructor rarely had to do so, indicating that students were correctly relating the questions.

Prerequisite Relations Quality

The reported number of 1,255 prerequisite relations reflects only those remained after students removed inappropriate relations based on feedback. Another indicator that students were drawing high-quality relations is that not all of the relations students created were to recently added questions. The average difference between creation times of pairs of directly related questions was 14 days and 12 hours on the Python map and 17 days and 10 hours on the HTML/CSS map. This indicates that students were considering many previous questions as candidates and integrating their knowledge, not just connecting each new question to the most recently studied topic.

EVALUATION OF INNOVATIVE DESIGN FEATURES

In this section, we assess the impact of particular features on student motivation to create high-quality questions and relations. We can not definitively tease apart the effects of different features, since we did not compare student actions when using alternative designs. Our primary source of evidence is qualitative. We observed behavior throughout. At the end of the study period, we collected open-ended individual reflections and conducted a retrospective group reflection session. For the individual reflections, we asked the students to fill out an anonymous Qualtrics survey. There were no incentives for completing the survey. There were 13 open-ended questions. Each prompt asked students about possible alternative designs for one feature that they had experienced. Some student responses reflected on the feature as they had experienced it while others reflected on the hypothetical alternative(s). Eleven of the nineteen students filled out the survey. Both the survey responses and transcripts from the retrospective group reflection are in the supplementary material. Student names are anonymized and referred to as "S#"[2].

We assess only those design features that were innovative in our learning process. Additional features described above, such as students voting and commenting on each other's contributions and coordinating activity through topic sprints, likely contributed to maintaining student motivation. However, they have been widely used in other systems, and evidence from our study about these features offers little additional insight.

Organized QMap of Questions

Students found QMaps, especially the requirement to find and draw prerequisite relations on the maps, to be helpful. S8 wrote: *"[it helps] because you can quickly locate similar questions."* S9 wrote: *"By seeing all of the connections and nodes, students are given a sense of pride and purpose and are motivated to collaborate to grow the map."*

Requiring students to place questions on the map did seem to enhance student learning. Early on, a few students posted high-quality questions they found from external sources such as previous exams or assignments, but they were not well placed on the map. When an instructor provided feedback about incorrect relations, in some cases, it turned out that the students did not understand the material in the questions they had posted. Later in the semester, there were very few poorly placed questions, even though the task of placing them got progressively more difficult as the map got larger. Students also found the requirement to draw connections between questions helpful for their learning. S5 wrote: *"Students have to read the questions that cover the more general topic to find where their more specific question on the same topic fits into the map, this allows ... to review others' questions and improve the more general question ..."* This suggests that the requirement of placing questions in the map acted as a desirable difficulty, forcing students to understand and integrate their knowledge of the concepts in each question that they added.

In the group reflections, most students found the node alignment, which was maintained manually by the instructor, to be crucial for ease of navigating as the map got larger. S5 wrote: *"When students move nodes wherever they like it creates a more messy concept map."* In addition, we were worried that disorientation would occur due to difficulties with asynchronous collaborative concept mapping reported in the past literature [10, 63]. However, we did not receive any complaints about being disoriented by changes in the graph, during the study, or in individual and group retrospective reflections.

Ownership of Questions is Identified

Most students favored having their work identified. S5 wrote: *"[it] holds the designer responsible for the quality of his question, it also allows the designer to take pride. ... his peers start to associate his name with high-quality work and give him more upvotes."* S7 wrote: *"I also put more effort into writing [any one] question so that my name wouldn't be associated with crappy ones."* Identifiability also enabled private feedback, whereas anonymous contributions would only allow for public feedback. S6 wrote: *"[having identifiability,] students can ... send constructive feedback to the designer both privately or publicly on slack."* S8 wrote: *"Improving others' questions was ... easier because you could connect with them individually on slack."* Conversely, S2 wrote: *"students might feel more comfortable about judging the quality of the question itself if the person behind it is removed."* S1 wrote: *"students may feel less afraid to post more advanced questions [if] their identity is not revealed."*.

Ownership Transfer with Justified Edits

Generally, students preferred to create new questions from scratch, despite the instructor's encouragement to revise others' questions. Students did make direct edits to 134 questions posted by other students, 11.61% of the total. In each of these cases, the instructor determined that the revised questions/relations were improvements on the original. In a few other cases, the revisions made the questions worse; the instructor reverted these. In addition to the 134 questions where students made direct edits of others' questions, they also provided constructive feedback on 218 questions, leaving it to the original owners to make edits.

In the individual reflections, S3 wrote: *"[it] was a great system to reward students who spent a majority of their time improving other's questions. Without this process, these students participation would have been under-represented."* S7 wrote: *"No ownership and realistically not many people would bother*

[2] This study was exempted under IRB HUM00151656 because the interventions fell within the range of accepted educational practices.

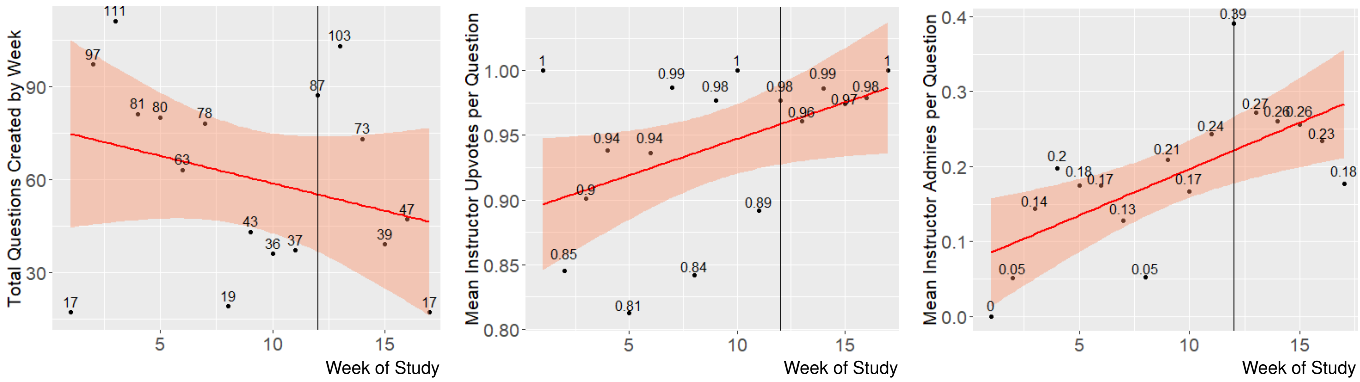

Figure 4. Questions & instructor votes by week. Vertical lines: the nudging week. Diagonal lines: fitted regressions. Confidence intervals are shaded.

fixing other people's questions. Partial ownership (partial reputation) and there might be some interest but making full question is still the better choice given." S11 wrote: *"[it motivates us] to take a look at past questions and constantly improve the map."* S9 opposed this feature: *"[students would] nit-pick through questions in order to change them and take ownership of them, rather than focusing on the big picture."* In the group reflections, the consensus was that because the original author of each question continued to be findable through the same slack search process that identifies the current owner, claiming was not too punitive.

Students also appreciated the requirement to post justifications for their modifications to the Slack channel. S6 wrote: *"[g]iving students reasons for any modifications can let students understand what they can do better... improving others' questions without explaining may kill students' motivation."* S8 wrote: *"To explicitly state why you modified each question allows you to thoroughly think through why the first question wasn't strong and what you could improve to make the question more applicable."* S2 wrote: *"original writers would improve their mistakes while the people who found the mistakes would fully understand and absorb the meaning of it."* Also, the requirement of justifying edits mitigates the concern of S9 above about a temptation to make unreasonably small improvements to claim the ownership of others' questions.

Visible Log of Activities

In addition to showing students a chronological log of all posts, Slack also notified students of new contributions. Students liked this feature. S5 wrote: *"[It] helps students keep track of modifications on each map and understand peer's reasoning for changing a question."* S1 wrote: *"Being able to see other students ... encourage[s] communication that is vital to the project. Students who can see the contributions of others will better contribute to the project themselves."* S9 wrote: *"This log ... [gives us] a better understanding of how to write a quality modification and see how questions related to the topic were previously improved."*

Instructor Assessment

Students overall appreciated the instructor assessments. In individual reflections, S9 wrote: *"the instructor will be able to survey the students' understanding of a topic in general, and will be able to help if the whole group is struggling with a certain topic, rather than students just feeling 'lost' when they don't understand something."* S7 wrote: *"If students are left governing themselves with self-interests not necessarily leaning towards learning the subject deeply, quality will fall."*

Role Modeling

In addition to the $1,154$ questions that the students generated, in our study, the instructor created a total of 161 and 184 questions about Python and HTML/CSS topics, respectively. Students found instructor participation helpful. S5 wrote: *"[It] fosters a strong community among the team members and makes the instructor feel much more approachable."* S2 wrote: *"It gave me a metric of what to aim for when making questions."*

However, for one affordance given to the students, claiming ownership of questions, some students argued that having the instructor model that behavior could negatively impact students' motivation. S9 wrote: *"Having the instructor just give feedback ... rather than claim ownership ... students will not feel that they are being stepped over when they do something wrong."* S2 wrote: *"simply giving feedback is obviously helpful ... students may learn more by being forced to fix the mistakes by themselves."* Following these negative reflections, we suggest that instructors who use our design should only give constructive feedback and not directly improve student-generated questions. This way, students would learn how to improve their questions and resolve misconceptions without being worried about losing credit for their contributions.

DISCUSSION

Overall, students engaged well with all three of the core activities: creating questions, creating relations, and providing feedback that led to improvements in the questions and relations. The quality of questions and relations was high and improved over time. Moreover, students maintained their motivation to participate throughout the summer without the external incentives of grades (with the exception of one student who earned one course-credit). Here, we place some of the design choices in a broader context.

A Focus on Relations

Many learning theorists have emphasized the value of assimilating new knowledge with existing knowledge, leading to what Ausubel calls meaningful learning. It is not obvious, however, how to organize pedagogy and assessment practices around this goal. Our approach is, in some sense, minimalist. The only expectation placed on students was when adding a new question, they had to identify at least one prerequisite, one question that someone should be able to answer before they try to answer the newly added question.

This minimal requirement turned out to be surprisingly powerful. It led students to notice and engage with content that others had posted and prevented them from adding duplicates or questions they did not understand. Even without any explicit injunction for students to connect new questions with past material, to improve meaningful learning, the mean difference between creation times of linked questions was more than two weeks.

Still, it would be interesting to explore more maximalist approaches, where the relations between concepts become an even larger focus of attention. One possibility within our framework of activities would be to encourage students to create multiple choice questions *about the relations* between concepts. For example, if there are existing questions about Python lists and tuples, one or more questions might test the understanding of how they are alike and different.

Scalability of Maps

All approaches to visual mapping must confront the challenge of scale. As maps get bigger, navigation and even orientation become more difficult. The general approaches for dealing with this involve clustering (putting related things spatially near each other), modularity (putting each cluster on a different map, with cross-map links treated differently), and abstraction (collapsing a node or region when zoomed out).

In our case, as in most approaches, spatial clustering emerged naturally as a result of placing nodes near linked nodes. We employed a form of abstraction by having collapsible nodes, with only question stems visible when closed. To help with orientation, we maintained two spatial conventions: the left-to-right ordering of prerequisite relationships and the grid alignment of fixed-width nodes. The fixed width and visual alignment make it harder for readers to depend for orientation on visual anchors in the form of nodes of unusual size or placement. However, the same-width nodes with grid alignment reduce visual complexity, and collapsible nodes allow for zooming out to get a bigger picture.

Different approaches to the challenge of scale in visual mapping may be more appropriate for learners at different levels. We found that a single QMap could be navigated even as it grew to include an entire semester's worth of programming concepts, the equivalent of an entire textbook. Novakian concept maps generally are limited to a smaller amount of material, requiring a division of the content into several smaller maps with limited ability to express links between them.

We speculate that QMaps may be helpful for high-achieving college-level students or professionals to visualize complex relationships, but the maps might be too challenging to navigate through for less advanced students who are not able to maintain a mental model of the overall map structure. Novakian concept mapping is likely to be a better choice for younger learners, especially when drawing maps using pen and paper.

Collective Editing

A fundamental challenge in any collective editing system, from Wikipedia to StackExchange to Quora to GitHub to Google docs, is to encourage actions that lead to the refinement of content over time, rather than just accumulation of additional material. This involves attracting the attention of multiple people to the same content, providing a mechanism for generating proposed changes, and a mechanism for deciding which changes will become part of the canonical version and the order in which alternative versions will be shown. Subtle variations of technical affordances and social norms can have a big impact on people's willingness to provide and refine content.

Like Wikipedia, our approach involved maintaining a single, current best version of each question, but we had a social convention whereby small changes were suggested to the current owner, with direct edits and claiming of ownership only for more major changes. This worked reasonably well, with ownership changes happening on about one-tenth of the questions with another two-tenths receiving suggestions for the original owners to make edits. We think there is room for improvement, however, and look forward to exploring alternatives in future work. We are particularly inspired by StackExchange, which has affordances for both editing existing answers to questions and proposing alternative answers, which are then sorted based on upvotes, and both affordances are widely used.

Student Privacy vs. Credit for Contributions

One of our concerns in the design process was whether to make the author of each question identified. Denny et al. [15] mentioned the anonymity of all activities in their study as a limitation and suggested more investigation. They were concerned that students might not wish to have their contribution identified, especially for those with lower achievements.

Our retrospective responses show that students found it very motivating and helpful that student names were associated with both the multiple-choice questions they added and the comments they wrote about each others' questions. Due to student privacy regulations, however, it might be difficult to implement our approach directly in credit-bearing courses where instructor feedback is used to determine grades. For example, imagine a *gameful learning* [1] approach where students earn points for each question that the instructor upvotes and certain point thresholds translate to different letter grades. Arguably, our setup with all instructor feedback public, would permit any student to determine any other student's grade.

The Role of the Instructor

Previous research by Denny et al. [15] did not explore the potential role of instructor in the process of voluntary question generation. They mentioned as one of the limitations of their design that receiving votes/comments only from peers may not be an adequate incentive to generate high-quality questions.

Somewhat ironically, we found that the instructor played a crucial role in our process, despite the fact that it was designed to be an extreme form of student-directed learning, with no deadline, grades, or required activities. Note that early on, it was challenging for some students to design high-quality questions and relations. It was very time-consuming for the instructor to give constructive feedback on the student-generated content. After a month, however, the comments, and questions and relations designed by the instructor helped students learn how to contribute high-quality content, and the instructor had to spend less time.

Perhaps teachers can breathe a sigh of relief that their role can still be important even in student-directed learning. In our study, students found the continuous assessment and constructive feedback on their questions by the instructor a crucial complement for peer-evaluation. Moreover, the students identified the questions and relations designed by the instructor as a guide for how to design questions and draw relations.

There are limits, however. It is possible for the instructor's participation to demotivate as well as motivate. In particular, we found that students disliked having the instructor directly improve and claim ownership of students' questions.

Limitations

In this study, we only tested the combination of all the discussed design features in the same study. Without A/B tests that turn on and off individual features, we cannot be entirely sure why the design worked. In addition, we only conducted a single trial, in learning programming languages with undergraduate students. It is necessary to run more studies in other disciplines, on different ranges of age, ethnicity, and nationality to learn about the adaptation and generalizability of the benefits of this learning process.

We only invited students who received *A* grades in at least one programming course. Presumably, only the most motivated students signed up to participate in a not-for-credit summer learning experience. Similarly, in another recent study of voluntary question generation, not for a grade [15], Denny reported that the students who used PeerWise were higher performers (i.e., earned higher course grades) and we speculate that those who chose to generate questions were even higher performing. This is a more general issue for self-directed student learning. It is necessary to conduct future studies on motivating low-performing students to engage in such question-generation activities. They may require very different motivators, and very different kinds of feedback than the high-performing students. It is an open question whether self-directed learning through question generation will work at all for populations beyond the highest-performing students.

Future Work

Using Draw.io to implement QMaps was problematic for students due to performance problems when each map grew after collaborating on it for more than a month. Also, going back and forth between Draw.io and Slack was confusing and annoying. This was mentioned by multiple students throughout the study. The fact that students were able to maintain motivation despite the sub-optimal tools is a further testament to the value of the learning process.

To better support QMapping in the future, we are now developing and integrating our two subsystems into a unified web application to make it easier for students and instructors to interact with the system and each other. As we do so, we will be guided by insights from this study. Particularly, we will try to automatically preserve layout constraints in the QMap, make it easier to identify the current owner of each question without searching the history log, and maintain a linear, chronological view of the activities in addition to the map.

We are also planning to integrate the question generation and mapping design with a retrieval practice system that we introduced in [78]. It will quiz students on the questions that they have collectively generated. In addition to the direct effect of reinforcing knowledge, improving students' long-term learning, we hope that it will also provide another prompt for students to provide feedback and lead to further improvements of the questions.

CONCLUSION

We report our design and deployment of an asynchronous, collaborative learning process. It combines multiple-choice question generation by students with a new knowledge visualization method, called "QMapping." One nice side effect of requiring students to place questions in a QMap is that it discourages entry of ill-understood questions copied from external sources. We identify several changes to conventional concept maps that make the maps more suitable for representing concepts in the form of multiple-choice questions.

We identify several innovative features that helped to sustain students' motivation for voluntary and collaborative creation and improvement of questions and relations. One theme is that identified rather than anonymous contribution was motivating for students. Another theme is the important role of the instructor, providing feedback, and serving as a role model. Two major challenges for the future are to make self-directed learning of this type work for lower-achieving students, and with less instructor effort.

ACKNOWLEDGMENT

We thank Jingcong Hu, Kallen Cohane, and Sydney Vollmar, three student participants who provided many ideas for improving the learning process.

REFERENCES

[1] Stephen J Aguilar, Caitlin Holman, and Barry J Fishman. 2018. Game-inspired design: Empirical evidence in support of gameful learning environments. *Games and Culture* 13, 1 (2018), 44–70.

[2] David G Ausubel. 1963. Cognitive Structure and the Facilitation of Meaningful Verbal Learning. *Journal of teacher education* 14, 2 (1963), 217–222.

[3] David P Ausubel. 1966. Ego development among segregated Negro children. In *Mental Health and Segregation*. Springer, 33–40.

[4] Simon P Bates, Ross K Galloway, and Karon L McBride. 2012. Student-generated content: Using PeerWise to enhance engagement and outcomes in

introductory physics courses. In *AIP Conference Proceedings*, Vol. 1413. AIP, 123–126.

[5] Anita Berglund. 2015. What's in a Word? Concept mapping: a graphical tool to reinforce learning of epidemiological concepts. *J Epidemiol Community Health* (2015), jech–2014.

[6] David Biggins, Emma J Crowley, Elvira Bolat, Mihai Dupac, and H Dogan. 2015. Using Peerwise to improve engagement and learning. The International Academic Forum (IAFOR).

[7] Danielle Blouin. 2018. Impact of interpersonal relations on learning and development of professional identity: A study of residents' perceptions. *Emergency Medicine Australasia* 30, 3 (2018), 398–405.

[8] Steven Bottomley and Paul Denny. 2011. A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. *Biochemistry and Molecular Biology Education* 39, 5 (2011), 352–361.

[9] David Boud. 2000. Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education* 22, 2 (2000), 151–167.

[10] Susan M Bridges, Esmonde F Corbet, and Lap Ki Chan. 2015. Designing problem-based curricula: The role of concept mapping in scaffolding learning for the health sciences. *Knowledge Management & E-Learning: An International Journal* (2015).

[11] Alberto J Cañas, Joseph D Novak, and Priit Reiska. 2015. How good is my concept map? Am I a good Cmapper? *Knowledge Management & E-Learning* 7, 1 (2015), 6.

[12] Alberto J Cañas, Priit Reiska, and Joseph D Novak. 2016. Is My Concept Map Large Enough?. In *International Conference on Concept Mapping*. Springer, 128–143.

[13] Paul Denny. 2013. The effect of virtual achievements on student engagement. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 763–772.

[14] Paul Denny. 2015. Generating practice questions as a preparation strategy for introductory programming exams. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. ACM, 278–283.

[15] Paul Denny, Fiona McDonald, Ruth Empson, Philip Kelly, and Andrew Petersen. 2018. Empirical support for a causal relationship between gamification and learning outcomes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 311.

[16] Paul Denny, Ewan Tempero, Dawn Garbett, and Andrew Petersen. 2017. Examining a Student-Generated Question Activity Using Random Topic Assignment. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, 146–151.

[17] Stephen W Draper. 2009. Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology* 40, 2 (2009), 285–293.

[18] Honglu Du, Mary Beth Rosson, and John M Carroll. 2012. Augmenting classroom participation through public digital backchannels. In *Proceedings of the 17th ACM international conference on Supporting group work*. ACM, 155–164.

[19] Denis Duret, Rob Christley, Paul Denny, and Avril Senior. 2018. Collaborative learning with PeerWise. *Research in Learning Technology* 26 (Jan. 2018). DOI: http://dx.doi.org/10.25304/rlt.v26.1979

[20] Martin R Fellenz. 2004. Using assessment to support higher level learning: the multiple choice item development assignment. *Assessment & Evaluation in Higher Education* 29, 6 (2004), 703–719.

[21] Denae Ford, Kristina Lustig, Jeremy Banks, and Chris Parnin. 2018. We don't do that here: How collaborative editing with mentors improves engagement in social q&a communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. ACM, 608.

[22] Lawrence T Frase and Barry J Schwartz. 1975. Effect of question production and answering on prose recall. *Journal of Educational Psychology* 67, 5 (1975), 628.

[23] JP Gerber, Ladd Wheeler, and Jerry Suls. 2018. A social comparison theory meta-analysis 60+ years on. *Psychological bulletin* 144, 2 (2018), 177.

[24] Elena L Glassman, Aaron Lin, Carrie J Cai, and Robert C Miller. 2016. Learnersourcing personalized hints. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1626–1636.

[25] Dale Hancock, Nicole Hare, Paul Denny, and Gareth Denyer. 2018. Improving large class performance and engagement through student-generated question banks. *Biochemistry and Molecular Biology Education* 46, 4 (2018), 306–317.

[26] Derek L Hansen. 2006. Knowledge sharing, maintenance, and use in online support communities. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1751–1754.

[27] Judy Hardy, Simon P Bates, Morag M Casey, Kyle W Galloway, Ross K Galloway, Alison E Kay, Peter Kirsop, and Heather A McQueen. 2014. Student-generated content: Enhancing learning through sharing multiple-choice questions. *International Journal of Science Education* 36, 13 (2014), 2180–2194.

[28] Stephen G Harkins and Jeffrey M Jackson. 1985. The role of evaluation in eliminating social loafing. *Personality and Social Psychology Bulletin* 11, 4 (1985), 457–465.

[29] ASC Hooper, Simon Jigwan Park, and Gemma Gerondis. 2011. Promoting student participation and collaborative learning in a large INFO 101 class: Student perceptions of PeerWise Web 2.0 technology.. In *Proceedings of Higher Education Research and Development Society of Australasia International Conference (HERDSA 2011)*.

[30] Kiri Hunter and Catherine Cook. 2018. Role-modelling and the hidden curriculum: New graduate nurses' professional socialisation. *Journal of clinical nursing* (2018).

[31] Sunyoung Joung and John M Keller. 2004. The Effects of High-Structure Cooperative versus Low-Structure Collaborative Design of Decision Change, Critical Thinking, and Interaction Pattern during Online Debates. *Association for Educational Communications and Technology* (2004).

[32] Steven J Karau and Kipling D Williams. 1993. Social loafing: A meta-analytic review and theoretical integration. *Journal of personality and social psychology* 65, 4 (1993), 681.

[33] John Keller and Katsuaki Suzuki. 2004. Learner motivation and e-learning design: A multinationally validated process. *Journal of educational Media* 29, 3 (2004), 229–239.

[34] DoHyoung Kim and Frank M Shipman III. 2012. Visualizing history to improve users' location and comprehension of collaborative work. In *Proceedings of the 17th ACM international conference on Supporting group work*. ACM, 11–20.

[35] Juho Kim, Philip J Guo, Carrie J Cai, Shang-Wen Daniel Li, Krzysztof Z Gajos, and Robert C Miller. 2014a. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 563–572.

[36] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J Guo, Robert C Miller, and Krzysztof Z Gajos. 2014b. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 4017–4026.

[37] Ian M Kinchin. 2016. *Visualising powerful knowledge to develop the expert student: A knowledge structures perspective on teaching and learning at university*. Springer.

[38] Srikanth Kolluru. 2012. An active-learning assignment requiring pharmacy students to write medicinal chemistry examination questions. *American Journal of Pharmaceutical Education* 76, 6 (2012), 112.

[39] George P Landow. 1989. The rhetoric of hypermedia: Some rules for authors. *Journal of Computing in Higher Education* 1, 1 (1989), 39–64.

[40] John O Ledyard. 1994. Public goods: A survey of experimental research. (1994).

[41] Mark R Lepper, David Greene, and Richard E Nisbett. 1973. Undermining children's intrinsic interest with extrinsic reward: A test of the" overjustification" hypothesis. *Journal of Personality and social Psychology* 28, 1 (1973), 129.

[42] Liang-Yi Li and Chin-Chung Tsai. 2017. Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. *Computers & Education* 114 (2017), 286–297.

[43] Ching Liu, Juho Kim, and Hao-Chuan Wang. 2018. ConceptScape: Collaborative Concept Mapping for Video Learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 387.

[44] Kathleen Luchini, Chris Quintana, and Elliot Soloway. 2004. Design guidelines for learner-centered handheld tools. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 135–142.

[45] Andrew Luxton-Reilly, Daniel Bertinshaw, Paul Denny, Beryl Plimmer, and Robert Sheehan. 2012. The impact of question generation activities on performance. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education*. ACM, 391–396.

[46] Lena Mamykina, Thomas N Smyth, Jill P Dimond, and Krzysztof Z Gajos. 2016. Learning from the crowd: Observational learning in crowdsourcing communities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2635–2644.

[47] HA McQueen, C Shields, DJ Finnegan, J Higham, and MW Simmen. 2014. PeerWise provides significant academic benefits to biological science students across diverse learning tasks, but with minimal instructor intervention. *Biochemistry and Molecular Biology Education* 42, 5 (2014), 371–381.

[48] Scott A Miller and Celia A Brownell. 1975. Peers, persuasion, and Piaget: Dyadic interaction between conservers and nonconservers. *Child Development* (1975), 992–997.

[49] David Nicol. 2007. E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and higher Education* 31, 1 (2007), 53–64.

[50] Joseph D Novak. 1990. Concept mapping: A useful tool for science education. *Journal of research in science teaching* 27, 10 (1990), 937–949.

[51] Joseph D Novak. 2002. Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science education* 86, 4 (2002), 548–571.

[52] Joseph D Novak. 2011. A theory of education: Meaningful learning underlies the constructive integration of thinking, feeling, and acting leading to empowerment for commitment and responsibility. *Meaningful Learning Review* 1, 2 (2011), 1–14.

[53] Joseph D Novak and Alberto J Cañas. 2008. The theory underlying concept maps and how to construct and use them. *Florida Institute for Human and Machine Cognition* (2008).

[54] Ferry Novliadi, Rika Eliana, and others. 2017. Reducing Social Loafing Tendency Through Achievement Motivation Training. *International Journal of Management Science and Business Administration* 4, 1 (2017), 15–18.

[55] Rianne Poot, Renske AM De Kleijn, Harold VM Van Rijen, and Jan van Tartwijk. 2017. Students generate items for an online formative assessment: Is it motivating? *Medical teacher* 39, 3 (2017), 315–320.

[56] Paul Resnick, Ko Kuwabara, Richard Zeckhauser, and Eric Friedman. 2000. Reputation systems. *Commun. ACM* 43, 12 (2000), 45–48.

[57] C Scott Rigby, Edward L Deci, Brian C Patrick, and Richard M Ryan. 1992. Beyond the intrinsic-extrinsic dichotomy: Self-determination in motivation and learning. *Motivation and emotion* 16, 3 (1992), 165–185.

[58] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.

[59] Richard Michael Ryan and Edward Lewis Deci. 2016. Facilitating and hindering motivation, learning, and well-being in schools: Research and observations from self-determination theory. *Handbook of motivation at school* 96 (2016).

[60] Andy Sharp and Angela Sutherland. 2007. Learning Gains... my (ARS) The Impact of Student Empowerment Using Audience Response Systems Technology on Knowledge Construction. In *Student Engagement and Assessment REAP International Online Conference on Assessment Design for Learner Responsibility, 29th-31st May*, Vol. 2007. 29–31.

[61] Yih-Chearng Shiue, Chao-Min Chiu, and Chen-Chi Chang. 2010. Exploring and mitigating social loafing in online communities. *Computers in Human Behavior* 26, 4 (2010), 768–777.

[62] Mel Silberman. 1996. *Active Learning: 101 Strategies To Teach Any Subject.* ERIC.

[63] Christina De Simone, Richard F Schmid, and Laura A McEwen. 2001. Supporting the learning process with collaborative concept mapping using computer-based communication tools and processes. *Educational research and evaluation* 7, 2-3 (2001), 263–283.

[64] Stephen Snow, Adriana Wilde, Paul Denny, and mc schraefel. 2019. A discursive question: Supporting student-authored multiple choice questions through peer-learning software in non-STEMM disciplines. *British Journal of Educational Technology* (2019).

[65] Yang Song, Zhewei Hu, and Edward F Gehringer. 2016. Who Took Peer Review Seriously: Another Perspective on Student-Generated Quizzes.. In *EDM (Workshops)*.

[66] Robert Sternszus, Yvonne Steinert, Farhan Bhanji, Sero Andonian, and Linda S Snell. 2018. Evaluating a novel resident role-modelling programme. *The clinical teacher* 15, 3 (2018), 252–257.

[67] Pei-Chen Sun, Ray J Tsai, Glenn Finger, Yueh-Yang Chen, and Dowming Yeh. 2008. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & education* 50, 4 (2008), 1183–1202.

[68] Chin-Chung Tsai, Sunny SJ Lin, and Shyan-Ming Yuan. 2002. Developing science activities through a networked peer assessment system. *Computers & Education* 38, 1-3 (2002), 241–252.

[69] Lise Vesterlund. 2016. Using experimental methods to understand why and how we give to charity. *Handbook of Experimental Economics* 2 (2016), 91–151.

[70] Dakuo Wang, Judith S Olson, Jingwen Zhang, Trung Nguyen, and Gary M Olson. 2015. DocuViz: visualizing collaborative writing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1865–1874.

[71] Xu Wang, Srinivasa Teja Talluri, Carolyn Rosé, and Kenneth Koedinger. 2019. UpGrade: Sourcing Student Open-Ended Solutions to Create Scalable Learning Opportunities. (2019).

[72] Sarah Weir, Juho Kim, Krzysztof Z Gajos, and Robert C Miller. 2015. Learnersourcing subgoal labels for how-to videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 405–416.

[73] James V Wertsch and Peeter Tulviste. 1992. LS Vygotsky and contemporary developmental psychology. *Developmental psychology* 28, 4 (1992), 548.

[74] Joseph Jay Williams, Juho Kim, Anna Rafferty, Samuel Maldonado, Krzysztof Z Gajos, Walter S Lasecki, and Neil Heffernan. 2016. Axis: Generating explanations at scale with learnersourcing and machine learning. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 379–388.

[75] Kipling D Williams and Steven J Karau. 1991. Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of personality and social psychology* 61, 4 (1991), 570.

[76] Chun-Ping Wu, Yu-Chi Chen, and Shu-ling Wu. 2018a. Development of a collaborative learning space for student generated-question strategy. In *Proceedings of the 6th International Conference on Information and Education Technology*. ACM, 102–105.

[77] Chun-Ping Wu, Jie Tu, and Shu-ling Wu. 2018b. Design a reflective mechanism to engage students in student-constructed test activity. In *Proceedings of the 6th International Conference on Information and Education Technology*. ACM, 106–110.

[78] Iman YeckehZaare, Paul Resnick, and Barbara Ericson. 2019. A Spaced, Interleaved Retrieval Practice Tool that is Motivating and Effective. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. ACM, 71–79.

[79] Xi Zhang, Shan Jiang, and Yihang Cheng. 2017. Inferring the Student Social Loafing State in Collaborative Learning with a Hidden Markov Model: A Case on Slack. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 149–152.