Crowd Coach: Peer Coaching for Crowd Workers' Skill Growth

CHUN-WEI CHIANG[1], HCI Lab, West Virginia University (WVU)[1], USA
ANNA KASUNIC[2], Carnegie Mellon University (CMU)[2], USA
SAIPH SAVAGE[1,2,3], Universidad Nacional Autonoma de Mexico (UNAM)[3], Mexico

Traditional employment usually provides mechanisms for workers to improve their skills to access better opportunities. However, crowd work platforms like Amazon Mechanical Turk (AMT) generally do not support skill development (i.e., becoming faster and better at work). While researchers have started to tackle this problem, most solutions are dependent on experts or requesters willing to help. However, requesters generally lack the necessary knowledge, and experts are rare and expensive. To further facilitate crowd workers' skill growth, we present Crowd Coach, a system that enables workers to receive peer coaching while on the job. We conduct a field experiment and real world deployment to study Crowd Coach in the wild. Hundreds of workers used Crowd Coach in a variety of tasks, including writing, doing surveys, and labeling images. We find that Crowd Coach enhances workers' speed without sacrificing their work quality, especially in audio transcription tasks. We posit that peer coaching systems hold potential for better supporting crowd workers' skill development while on the job. We finish with design implications from our research.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**;

Additional Key Words and Phrases: Crowdsourcing, Amazon Mechanical Turk, Worker training, Peer review, Peer advice, Future of work

ACM Reference Format:
Chun-Wei Chiang[1], Anna Kasunic[2], and Saiph Savage[1,2,3]. 2018. Crowd Coach:, Peer Coaching for Crowd Workers' Skill Growth. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW, Article 37 (November 2018), 17 pages. https://doi.org/10.1145/3274306

1 INTRODUCTION

In traditional employment, workers usually have the opportunity to grow their skills over time [9, 30, 39, 85]. Work satisfaction theories have stressed the importance of fostering skill development in work environments to increase workers' motivations and contentment [9, 18, 68, 71]. This desire to develop one's skills is also present in crowd workers [46]. However, crowd markets in general have not been designed for skill growth [8, 26, 76, 83]. Consequently, crowd workers who wish to extend their skills must explore ways to train themselves outside crowdsourcing platforms [48]. However, given the low pay of crowd work [6, 29, 34, 66, 77], requiring workers to use additional time and money for skill development is impractical [1, 47, 69, 81]. To address this issue, scholars have recently started to explore models to enable skill growth while doing crowd work [22, 26, 27]. However, these

Authors' addresses: Chun-Wei Chiang[1], HCI Lab, West Virginia University (WVU)[1], USA, cc0051@mix.wvu.edu; Anna Kasunic[2], Carnegie Mellon University (CMU)[2], USA, akasunic@andrew.cmu.edu; Saiph Savage[1,2,3], Universidad Nacional Autonoma de Mexico (UNAM)[3], Mexico, saiph.savage@mail.wvu.edu.

2573-0142/2018/11-ART37 $15.00
https://doi.org/10.1145/3274306

models have relied heavily on requesters, whose time is limited, or domain experts, who are expensive [76]. Requesters also have insufficient knowledge or motivation for helping workers [27, 43, 51]. Consequently, these models are usually scarce and do not always address workers' needs [74, 75].

To enable skill development in crowd workers without requiring experts, we introduce the system Crowd Coach: a Chrome plugin that provides workers with short advice from peers while working on AMT. Crowd Coach provides two primary mechanisms to support skill development on crowd markets: (1) Micro-Coaching and (2) Selective Coaching. Figure 1 presents an overview of our system. The Micro-Coaching interface encourages workers to provide short snippets of advice (coaching snippets) to each other while doing crowd work. This approach draws on "Twitch crowdsourcing" research [80] that focuses on enabling people to do micro-tasks rapidly and without them being a person's main activity. Based on these ideas, we specifically limit the coaching that workers provide to a length of 100 characters. This empowers workers to peer-coach while still doing their main job on AMT. In parallel, our Selective Coaching mechanism helps workers to follow the coaching that might best help them to develop their skills. In specific, we focus on developing workers' ability to do specific tasks better over time [79], which is typically measured in terms of the quality of work produced and the amount of time to complete the task [3, 45].

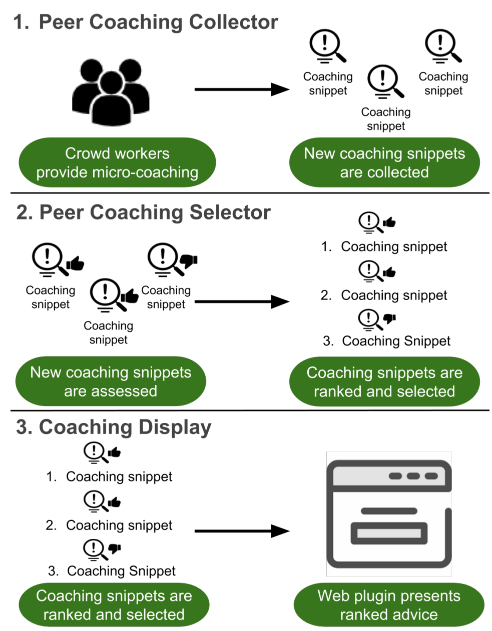

Fig. 1. Overview of Crowd Coach's functionality.

In this paper we contribute: 1) a system supporting crowd workers' skill development via peer coaching; 2) a mechanism to enable providing and receiving peer coaching while doing crowd work; 3) a field experiment demonstrating that Crowd Coach facilitates peer-coaching and benefits the crowd workers receiving the coaching; 4) a real world deployment showcasing how workers use Crowd Coach in the wild to better uncover the system's benefits and drawbacks. We believe Crowd Coach's method of using short, selective peer coaching has great potential for transforming the ways in which crowd workers can develop their skills while on the job.

2 RELATED WORK

Crowd Coach's system design is informed by four main research areas: 1) Skill development in crowd work; 2) Tools for improving crowdsourcing platforms; 3) Online Peer to Peer Support and Collaborations; 4) Peer learning.

1. Skill Development in Crowd Work. Both researchers and practitioners still struggle with enabling skill development on crowdsourcing platforms [23, 77]. Enabling skill develop is important for several reasons. First, skill improvement enhances performance in crowd work [48]. Complex tasks also need skilled workers to complete them [64]. Having more specialized workers can facilitate the completion of more work and facilitate reaching more complex goals [20]. Second, skill development can help crowd workers to finish their tasks faster [8]. This can potentially aid crowd workers to obtain a higher hourly pay, as payment on AMT is typically based on workload (number of tasks completed) [16, 42]. Third, "learning new skills" is one of the main motivators that crowd workers have for joining the crowd market [46]. Crowd workers expect that their experiences on crowdsourcing platforms will benefit their career development [11]. Nevertheless, most crowd workers lack the opportunity for improving or advancing their careers [76].

To address this problem, researchers have developed tools to facilitate skill development in crowd work. Atelier for Upwork [76] provided a mentorship model where experts helped crowd workers learn new skills; LevelUp for Photoshop [26] integrated tutorials designed by experts to teach people designer skills while completing basic photograph improvement tasks. However, most of these tools require experts' time to help new workers, either via mentoring or by developing educational material such as tutorials. Yet, experts' time is limited and expensive [76].

To reduce the dependency on experts, very recent research has started to explore tools that incorporate ideas from peer learning. Peer learning is a concept from cognitive psychology where students have the power to teach and learn from each other [10]. Prior work has uncovered that students who participate in peer learning perform better than students working on their own [56], and were even better than students instructed by experts [38]. This learning method also appears to be suitable for crowd workers; researchers have found that crowd workers enhance the quality of their work when they reviewed and provided feedback to the labor of other crowd workers [22, 27, 88]. However, these approaches assume that requesters will be willing to take the time to redesign their tasks to incorporate and facilitate the peer learning model. Under the current dynamics of AMT, it is likely that requesters lack: a) the knowledge on how to effectively design tasks and workflows to help workers [51]; b) the time and incentives to want to improve tasks and worker satisfaction [43]. Moreover, these studies made a trade-off between improving workers' labor quality and helping workers to reduce their completion time [51]. Prior research focused on helping workers produce higher quality work but with the cost of increasing their completion times. For instance, very recent investigations explored workflows where novice workers reviewed each other's work by writing long writeups about each other's labor (800 to 1,000 characters long) [27] or engaging in detailed discussions with each other [88]. While such workflows did help crowd workers to improve their performance, some workers also had to invest 25% more time in writing, reading and digesting each other's advice. Most crowd workers might not have the time to engage in such lengthy reviewing activity in addition to their main work. Morever, most workers are more interested in enhancing their skills to ultimately improve their salary than in increasing their work quality [57, 58]. As explained above, increasing crowd workers' completion time per task can decrease workers' hourly wage. We believe that crowd workers might also be able to improve while on the job without sacrificing quality over completion time by providing quick, but selective peer advice. We consider this selective advice to be similar to coaching that is usually very direct and precise. Kittur et al. [48] envisioned that crowd markets could be a direct medium for skill development. Crowd Coach builds on this idea by helping workers to develop their skills without requiring experts to invest their time or requesters to redesign the tasks. Instead, Crowd Coach uses peer micro-advice to guide crowd workers to improve their performance, especially in working time without sacrificing quality. Crowd Coach also provides a quality assessment loop that dynamically finds high-quality peer advice (selective advice).

2. Tools for Improving Crowdsourcing Platforms. Due to the inequalities between requesters and workers in AMT, workers usually obtain unfair treatment [43, 44]. Researchers and practitioners have been exploring different browser extensions to improve the working conditions of crowd workers and reduce the inequalities. The web plugin of Turkopticon allows crowd workers to evaluate requesters after finishing their tasks [43]. Turkopticon has become a useful tool for reducing the information asymmetry gap between workers and requesters (on AMT only requesters could officially rate workers). Other tools have explored enabling workers to help each other find higher hourly paying jobs to offset the fact that most crowdsourcing platforms do not provide an estimated working time of each task [16]. TurkBench proposed a market visualization tool to help crowd workers to better manage their tasks [33]. These novel tools have improved the working environment of crowd workers. In a similar fashion, Crowd Coach aims to improve crowd workers' labor conditions by facilitating on-the-job skill development.

3. Online Peer to Peer Support and Collaborations. Recently, we have seen the emergence of systems that coordinate peers of online strangers to share useful information with each other [62]. Several human computation workflows have successfully driven strangers to share their knowledge to help others learn [82]. These studies have found that online strangers can indeed provide quality information [63], even when asked by bots [72]. Related work has also investigated how receiving feedback or advice from peers affects the quality of online work produced in peer production communities [25, 37], such as Wikipedia [5, 73, 89]. We motivate the design of Crowd Coach on some of the key findings of this previous research: it is possible to drive online strangers to provide useful information for others [63, 86], and receiving peer advice can enhance a person's work [89]. We hypothesize that if we make it simple enough to provide peer advice, we will be able to orchestrate the coaching of crowd workers for improving their skills while on the job.

4. Peer Learning. In the design of Crowd Coach we posit that we can coordinate crowd workers to coach others via peer advice. Prior work has showed that peer advice can help students to improve their grades [32, 52]. Usually, this is because peer advice can help students to: (1) focus on the aspects of their work that will lead to the highest grade increase [49]; (2) better identify why they have poor performance [4]; or (3) encourage students to have higher work standards after seeing the great strides made by their peers [36, 55]. More recent research has designed and studied peer learning systems that maximize these advantages. Kulkarni et al. [53] studied systems to provide rapid peer advice to MOOC learners. Yoon et al [87] studied multi-modal peer advice systems. Yet most related work has studied peer learning within formal educational settings [17, 52, 78]. We expand this research and study integrating peer advice in informal learning environments, specifically crowd markets. Our work also investigates the potential of enabling peer feedback while the individuals are doing another main activity, specifically crowd work.

3 CROWD COACH

Motivated by the challenges of enabling crowd workers' skill development without depending on experts or requesters, our research translates peer learning models into the general design mechanism of "crowd coaching." In crowd coaching, workers can provide and access peer advice (coaching snippets) to improve their skills. While there are many ways workers could give and obtain peer coaching, we focus on coaching that could take place while on the job. We considered it was important to minimize the time that crowd workers spend outside AMT to reduce the instances where workers are not receiving wages. We therefore envisioned a web plugin solution that would allow people to continue on AMT and avoid disrupting their normal work routines. We also considered that the coaching would not be the primary task that individuals are doing. Therefore, we design both the activities of coaching and receiving coaching to be lightweight and avoid distracting people. We integrate into our design ideas from "Twitch crowdsourcing" research [80] where people do micro-tasks without disrupting their main activity. For this purpose, we specifically frame Crowd Coach's design around: (i) Availability: ability to evoke peer coaching with a click; (ii) Low Cognitive Load: allow crowd workers to coach and obtain coaching without the activity becoming a distraction to their main job. Given crowd workers' labor conditions [34], it was also important to integrate into our design: (iii) Paid Training: empower workers to personally improve their skills while earning money from their main job.

To enable these points, Crowd Coach has a: 1) Peer Coaching Collector, 2) Peer Coaching Selector (to select the coaching that is most useful), and 3) Coaching Display (to present useful coaching). Figure 1 presents an overview of Crowd Coach and Figure 2 showcases an overview of our system's interface.

1. Peer Coaching Collector. The Peer Coaching Collector enables workers to coach others by inputting advice on how to improve at particular HITs. The collector lives as a plugin on AMT where with a click, workers can become coaches to their peers and provide short coaching snippets on how to become faster and maintain good work quality for specific tasks (especially for the tasks that workers, i.e., coaches, are currently doing within their main job). Unlike previous research [27] that leads workers to give lengthy information, Crowd Coach encourages micro-advice. We view this micro-advice as similar to what a coach would briefly yell to players during a training to help them personally improve. We thus call this advice "coaching snippets". For users of the plugin (Google Chrome extension), Crowd Coach shows a small "provide tip" button on the AMT status bar. Upon clicking the button, workers see a small pop-up window where they can provide their coaching snippets. Notice that this facilitates our design principle of "availability to coach." We also wanted to limit the cognitive load that coaching imposes on already occupied workers. We thus limit the length of people's coaching to 100 characters and try to make the coaching as simple as possible (we set the characters limit to 100 characters through trial and error). In the pop-up window, workers simply select the type of tasks for which their coaching snippets is relevant and then type their coaching snippet. This design enables us to match coaching to particular tasks without imposing a large amount of new labor on workers. Notice that although AMT tasks do provide description and titles, and we could use topic modeling or NLP techniques to categorize the tasks, prior work has shown that such written information is not always the most relevant to classify tasks [34]. We therefore focused our efforts towards a crowdsourced solution. Based on prior work [34, 70], Crowd Coach considers that 8 main types of tasks exist on AMT and asks workers to classify tasks into either: Audio transcription, Categorization, Data Collection, Image Transcription, Image Tagging / Labeling, Surveys, Writing and Other.

2. Peer Coaching Selector. For each type of AMT task, the Peer Coaching Collector returns a long queue of coaching snippets. However, some coaching snippets may not be that useful for workers' personal growth. Especially with a large number of snippets, relevant "advice gems" might get lost in the muck. To overcome this issue, we have a Peer Coaching Selective module that focuses on identifying the best coaching snippets for a task.

The module first asks workers to micro-assess a coaching snippet via upvotes or downvotes. (see Figure 2). Crowd Coach tries for workers to micro-assess coaching related to tasks that workers are currently doing. When a coaching snippet gets assessed, it will win or lose "credits" (depending on whether it was upvoted or downvoted). These credits are pro-rated based on the reputation of the worker assessing. The system calculates the reputation of workers based on: a) their experiences on AMT (number of tasks completed); and b) how similar their micro-assessment is to that of other workers. If a person's micro-assessment deviates too much from what others have input, we consider it as an "alternative" assessment. Using this metric we classify micro-assessments as either "mainstream" or "alternative," and rank mainstream coaching based on peer ratings. By default, Crowd Coach presents first highly ranked mainstream coaching snippets to workers. Through its voting mechanism, Crowd Coach can thereby select coaching that peers found useful.

3. Coaching Display. This component focuses on presenting coaching that will help workers to develop their personal skills while on the job. For a given task, the Coaching Display presents to workers four associated coaching snippets that were ranked highest on the list. If workers want to read more coaching snippets, they can click the left or right button to view more. To ensure that new coaching snippets have the opportunity to be evaluated, Crowd Coach always mixes new

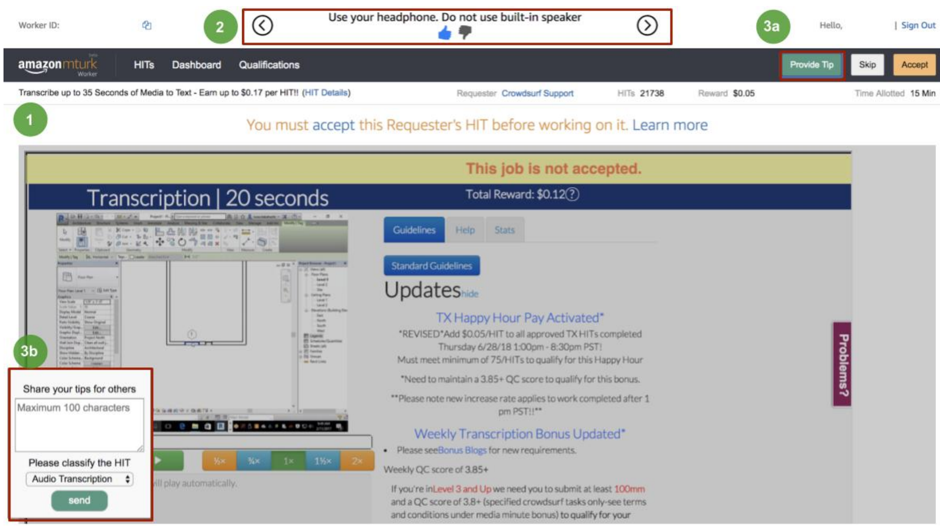

Fig. 2. Screenshot of our system in action. Crowd Coach enables workers to develop their skills while on the job by: (1) integrating directly into AMT hits; (2) presenting selected coaching snippets to workers; and (3) allowing workers to easily coach.

coaching snippets that needs micro-assessments into the list of high ranking coaching snippets. This dynamic helps to make sure that the new coaching snippets have the opportunity to be voted. Notice that workers can also choose to add their own coaching for a given task. Crowd Coach keeps track of the coaching that has already been shared with workers to avoid redundancy. Figure 2 presents how the coaching is displayed to workers and how workers coach peers.

4 EVALUATION

This paper hypothesizes that receiving selected peer coaching can help workers get started in building their skills. Similar to [27], we measure skill growth in terms of an increase in workers' speed and labor quality. To test this hypothesis and to understand the type of work that is well or poorly supported by Crowd Coach in the wild, we conducted: (1) a controlled field experiment; and a (2) real world deployment of Crowd Coach. The controlled field experiment allows us to study how Crowd Coach might facilitate crowd workers' skill growth. The deployment allows us to investigate natural usages of Crowd Coach to probe the strengths and weaknesses of our system.

4.1 Controlled Field Experiment

The goal of our field experiment was to compare our peer coaching approach with other conditions to evaluate its effectiveness in developing crowd workers' own skills. We considered 3 conditions: 1) workers do tasks without receiving any type of peer coaching [control condition]; 2) workers do tasks while receiving random coaching snippets[random snippets condition]; 3) workers do tasks while receiving selected peer coaching [Crowd Coach condition].

Given that for our experiment we needed to measure participants' work quality, we focused on skill development for labor that was not open-ended and whose quality we could more easily measure. We specifically focus on audio transcription tasks whose quality is directly measured with people's transcription accuracy. Not only are audio transcription tasks one of the most common AMT tasks [24], but in addition, becoming good at audio transcription can substantially increase a person's wages. Transcribers typically earn $0.01-0.02 USD per sentence they transcribe [65], which could potentially translate to high wages if a worker is fast (and accurate) enough. Specializing in audio transcription could allow crowd workers to command higher wages, as audio transcription is in high demand. Written records of court proceedings and captions for live television events, such as the news, sports, and political speeches all require real-time audio transcription. Audio transcription skills are thereby highly specialized, highly valued, and well paid, earning up to $300 per hour outside AMT. Building audio transcription skills on AMT could thereby help crowd workers expand their horizons and increase their earnings. Given all of this, we considered it valuable to study Crowd Coach's effectiveness in improving workers' audio transcription skills, especially for the case of novice workers. We considered that novices were the ones who could benefit the most from systems like Crowd Coach as it can be difficult to learn the ropes of the AMT ecosystem while also developing skills. Our field experiment thus focuses on investigating whether Crowd Coach helps novice crowd workers improve their audio transcription skills.

4.1.1 Field Experiment: Method. We studied novices' completion time and work quality for 3 different audio transcription tasks under one of our 3 study conditions. We considered that novice workers were both workers who were new to AMT and inexperienced in audio transcription tasks. We first recruited crowd workers who had completed less than 100 HITs on AMT. Next, we identified which of these workers were also novices in audio transcription. For this purpose, we asked potential participants to do a pre-test (which consisted of completing real world audio transcription tasks). We included in our study only the workers who finished such tasks in a similar time and with similar work quality (accuracy). We recruited a total of 90 novice AMT workers, and randomly divided 30 participants into one of our 3 experimental conditions. Participants in each condition were assigned the same audio transcription tasks with the same order. We collected the coaching snippets and micro-assessments of the coaching snippets before the experiment. Workers for each task were shown the same random coaching snippets or the same selected coaching snippets (depending on their condition.) Workers were paid $0.6 for completing an audio transcription tasks ($1.8 in total). We paid workers $0.6 when they accepted and worked on the first audio transcription task, and gave another $0.6 as bonus when they completed one more task. Workers could dropout whenever they wanted and get paid for the work they had finished. Tasks were sourced from real world audio transcription HITs on AMT and had similar difficulty: participants had to transcribe around 28 seconds of audio with similar levels of background noise, and with an average speaking rate of 165 word-per-second. To better trace participants' performance from task to task, we had participants interact with an AMT-like website we built. The website recorded workers' retention rate, completion time and accuracy for each task. To measure time to complete a task, we measured the time when a worker first accessed the task as the start time, and the time when workers submitted their labor (transcription) as their finish time. To study accuracy, we calculated the word error rate (WER) produced by each worker for each transcription, a commonly used metric to assess performance in audio transcription [7]. Participants also completed a survey about their experiences in each condition.

4.1.2 Field Experiment: Result. During the study period, novice workers completed a total of 253 tasks across all 3 conditions. Some novice workers (13) decided to not complete all the tasks and dropped out. 26 workers remained in the control group finishing all three tasks, 26 workers in the random coaching group, and 25 workers in the Crowd Coach group. The retention rates per condition were similar ($\chi^2(2) = 0.03$, p = 0.987). Figure 3 presents the task accuracy (work quality) and the completion time of workers who completed all three tasks under different conditions.

We used an one-way MANOVA to compare whether workers' differences in task accuracy and completion time were significant across conditions. Notice that the one-way MANOVA helps us to

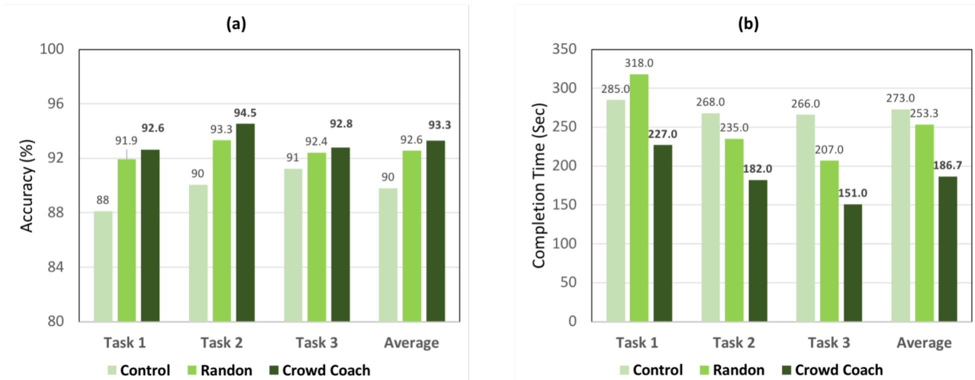

Fig. 3. (a) Workers' accuracy for each audio transcription task when working with the different coaching methods. (b) Workers' completion time for each task when working with the different coaching methods. Crowd Coach training outperformed other conditions in both accuracy and completion time.

determine whether there are any differences between independent groups (conditions) with more than one continuous dependent variables. In this case we have 3 independent groups (conditions) and 2 continuous dependent variables (task accuracy and completion time.) The one-way MANOVA also helps us to study the two dependent variables at the same time. This is important as there might be a trade off between becoming faster at audio transcription tasks and workers' task accuracy.

The one-way MANOVA showed that there were significant differences in task accuracy and completion time across the 3 conditions ($F(2,74) = 17.53$, $p < 0.0001$). Given these significant differences that we discovered, we ran an univariate ANOVA per dependent variable (time and accuracy). We found that the difference were significant in completion time ($F(2,74)= 54.18$, $p < 0.0001$), but not significant in task accuracy ($F(2,74)= 2.47$, $p = .09$). Next, we conducted a Tukey test as post hoc analysis to understand more deeply the difference in time completion between conditions. We found that the competition time in the Crowd Coach condition ($M = 184.1$, $SD = 12.36$) was significantly less than the competition time in the control condition ($M = 262.79$, $SD = 37.38$) at $p < 0.0001$ and the competition time in the random coaching condition ($M = 284.21$, $SD = 46.44$) at $p < 0.0001$.

Overall, our field experiment indicates that workers exposed to selected coaching snippets (Crowd Coach) were faster without sacrificing accuracy than workers without such coaching.

4.2 Deployment Study

We conducted a real world deployment of Crowd Coach to understand the type of work that is well or poorly supported by our system. We launched Crowd Coach and studied its use from June 25 to July 7th 2018. Similar to [16, 76], we paid workers $0.4 to install Crowd Coach in our deployment.

4.2.1 Deployment: Result. Crowd Coach was installed by 179 workers, who successfully coached or received coaching for tens of tasks. 86% of our participants were active users of the system (i.e., they either coached or micro-assessed the coaching); the rest used Crowd Coach more passively (they installed and potentially read the coaching.) 96 workers provided 363 coaching snippets, and 146 workers provided 1,401 micro-assessments to these coaching snippets. The median number of times workers coached each other was 2 and the worker who coached the most did it 32 times.

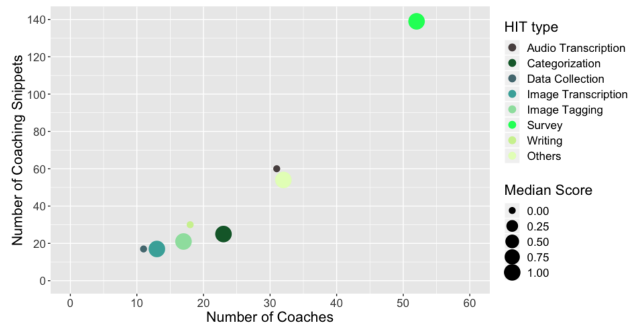

Fig. 4. Overview of the total number of coaches, coach snippets generated for each HIT type and the median score the coaching received.

Figure 4 and Table 1 present an overview of how workers used Crowd Coach across different types of tasks. Each point represents a type of task for which workers used Crowd Coach. The X axis represents the total number of coaches who participated in a particular type of task. The Y axis represents the total number of coaching snippets generated for each task type. The size of each point is proportional to the average assessment score (upvotes minus downvotes) of the coaching snippets associated with a particular type of task. From Figure 4 we observe that most coaches and coaching activity went into surveys and audio transcription tasks. Workers created a total of 139 coaching snippets for surveys, and 60 for audio transcription. We might be observing this preference for coaching surveys because they are one of the most common tasks on AMT and therefore workers are likely to have more experience with them. From Figure 4 we also observe that all types of tasks had at least ten different coaches, and at least 20 associated coaching snippets that were in general rated positively by workers (the median score was above zero.)

Workers overall tended to create different coaching snippets each time. Only 2% of all coaching snippets were repeated verbatim. We observed that when workers shared the exact same coaching snippets, they usually tended to change the categorization of the coaching snippets. For instance, first they posted the coaching snippets in audio transcription tasks and then in surveys. In these cases, workers seemed to share "general advice" that they considered was relevant across tasks. An example of such coaching: *"keep a list of requesters that you like completing hits for."* In future iterations of Crowd Coach, we are considering enabling workers' to state when their coaching is general and applicable across AMT tasks.

From Table 1 we observe that the coaching that received the highest maximum upvotes came from surveys, audio transcription, or tasks in the "other" category. Upon inspection, we identified that the coaching that workers upvoted the most tended to advise using smart interfaces to enhance one's work. For instance, one of the most upvoted coaching snippets involved advising workers to use an intelligent Chrome plugin: *"Use a multi-highlite extension for chrome to hilite words that you need to make sure you see in a survey."*. Workers also tended to upvote coaching that shared best practices for working with specific requesters. The following was also one of the most upvoted coaching snippets: *"With Stanford surveys remember to read carefully."* The coaching that workers tended to downvote the most was coaching that appeared to have been categorized incorrectly. For instance, the following coaching was the most downvoted: *"Some of these survey sites are legitimate but have either low pay for the work or only pay you in gift cards instead of cash."* The coaching

HIT Type	Coaching Snippets	Micro-Assessments	Micro-Assessments Score
Audio Transcription	Total=60; Per worker: max =10; median =1	Total=176; Per worker: max =10; median =2	Per task: min=-2; max =43; median =0
Categorization	Total=25; Per worker: max =3; median =1	Total=114; Per worker: max =8; median =2	Per task: min=-1; max =25; median =1
Data Collection	Total=17; Per worker: max =4; median =1	Total=81; Per worker: max =4; median =3	Per task: min=-2; max =22; median =0
Image Transcription	Total=17; Per worker: max =3; median =1	Total=62; Per worker: max = 6; median =2	Per task: min=0; max =16; median =1
Image Tagging	Total=21; Per worker: max =3; median =1	Total=84; Per worker: max =7; median =3	Per task: min=-1; max =16; median =1
Survey	Total=139; Per worker: max =21; median =2	Total=406; Per worker: max =28; median =2	Per task: min=-2; max =62; median =1
Writing	Total=30; Per worker: max =3; median =1	Total=114; Per worker: max =12; median =3	Per task: min=-3; max =18; median =0
Others	Total=54; Per worker: max =6; median =1	Total=364; Per worker: max =19; median =3	Per task: min=-3; max =61; median =1

Table 1. Overview of the results from our deployment. Workers used Crowd Coach for a wide range of tasks. There was a tendency to use our system to coach especially survey and audio transcription tasks.

clearly seems to be for surveys tasks. Yet the coach said it was for audio transcription tasks. From Table 1 we also observe that the coaching with most downvotes involved the "other" category which likely held a wide range of tasks. It is therefore probable that not all of the coaching in this category was relevant to workers. In the future we will explore allowing workers to correct the categories of coaching snippets. Coaching snippets that were also too specific tended to obtain the most downvotes. For instance, the following coaching snippet was one of the most downvoted and appears to be relevant just to the particular task the worker was doing:*"with this count on no bonus. It is a game of chance"*

We believe that users' opinion and feedback are the foundation to design future iterations of Crowd Coach. We designed a post-deployment survey to collect opinions from real users, and posted the survey as a $0.6 HIT on AMT available only to the workers who participated in our deployment. A total of 114 workers completed the survey, indicating a response rate of 64%, and the median time to complete the survey was 8 minutes. Our survey revealed that workers overall found Crowd Coach useful (on a scale of 1-5, median =4, mode = 4). 42% of the participants expressed that they felt they personally benefited greatly from having access to the coaching from others. In what follows, we present more details of the experiences with Crowd Coach that workers highlighted in the survey. We label worker participants as #W.

Being Coached Facilitates Exploring New Work Areas. Workers expressed that they used Crowd Coach to get a better sense of the experiences they would have if they did particular types of HITs. Being able to access such information seemed to help workers decide if they should venture into new work areas: *"[What I liked the most about Crowd Coach is] receiving tips about tasks you hadn't performed yet, helping you decide if you should take them."* (#W21).

Coaching is a Social Good Activity to Help the Personal Growth of Peers. Workers expressed that one of their main motivators for being a coach on our system was that they could help the personal growth of their peers: *"Knowing that there's someone out there struggling to learn to ropes just like me and I can help them is what got me to keep providing tips when I can"* (#W63).

Coaching Brings the Best of External Sites into AMT. Workers seemed to value that Crowd Coach allows them to stay on AMT while receiving important information they usually only obtained externally. *"I think this is a very useful tool. In theory, it's faster and more useful than relying on an external site (like Turkopticon) to peruse the ratings of requesters."* (#W89). Most plugins for crowd work still require workers to spend significant time outside AMT. For example, workers can rate requesters on Turkopticon by giving scores from 1 to 5. But, if workers want to know more details about requesters, they have to go to an external website. Workers liked that this plugin allowed them to have conversation type interacts with their peers without leaving AMT: *"I like the* *idea of interaction with fellow Turkers on the AMT site itself— very similar to MTurk Suite, but with actual conversations. It lets you interact with people."* (#W102).

Struggle Between Too General or Too Specific Coaching. Some workers felt that the system should guide coaching that was more specific rather than general: *"Make sure people are commenting on a particular hit and not just offering general advice ... like "getting a laptop", or that 'Wharton is usually a good one', I saw those on a lot of hits"* (#W6). While other workers argued that the coaching should be more general so that novice workers could learn from them: *"[...] I ended up leaving a couple of general tips–primarily useful for newbies."* (#W102).

5 DISCUSSION AND FUTURE WORK

We conducted a field experiment and a real world deployment to study different aspects of how Crowd Coach facilitated workers' personal skill development. Our field study experiment suggested that peer coaching can help novice crowd workers to improve in audio transcription tasks their speed while maintaining work quality. We believe that systems like Crowd Coach can likely empower crowd workers to increase their hourly pay as it could help workers to do more HITs per hour without sacrificing quality.

It is expected that in the future, crowd markets will become employment hubs where increasingly large numbers of workers with varying expertise and skill levels compete for employment and contribute to projects [48]. It therefore becomes important to envision mechanisms that help the workers of these platforms to continuously grow to obtain better opportunities [23].

Our field experiment suggested that we can use peer coaching to address the problem of facilitating personal skill development on crowd markets without requiring the insertion of experts. We also showed that for one of the most valued tasks on crowd markets— audio transcription tasks— even short coaching snippets that do not overwhelmingly distract from the task at hand can start to improve workers' speed. In the future, we would also like to explore using Crowd Coach to improve workers' personal skills for more complex professions such as CTOs, managers, or computer engineers. We also plan to explore the benefits of these type of crowd coach systems to facilitate skill development in rural communities where experts might be more scarce [2, 21].

Research has demonstrated that peers can provide advice that is just as effective as advice from experts [61, 67]. However, as with most pedagogy, the effects are not always consistent. Our field experiment and deployment can help researchers to better understand how peer advice plays out within the informal learning environment of crowd markets. Future work in this area could investigate more about how within crowd markets the coaching of peers differs from that of expert crowd workers

Prior research has identified that the order of micro-tasks impacts performance [15]. For instance, having spaced repetitions usually impacts how many words a person can learn on their own for vocabulary acquisition [31]; or mixing tasks with varied difficulty and similarity type impacts learning [50]. In the future, we plan to explore mixing peer coaching with the automatic generation of "lists of tasks to do." Helping workers to select the tasks that might be most beneficial for them while pairing them with the best peer-coaching might enable workers to grow even more.

Crowd Coach provides selected coaching snippets to novice workers for improvement. However, being exposed to peer coaching could also limit or shutdown workers' thought process and ideas on how to do a specific task. It is possible this could lead workers to believe that their approach is not best, as it is different than what others recommend, thereby discouraging workers from using creative approaches to solve tasks. In the future, we are interested in exploring ways to enable coaching that helps workers to improve, but also promotes workers' initiative and creativity [19]. We are interested in exploring different reward schemes in this space. Similar to [53], we are considering reward schemes used in domains like design where rapid iterations are prized [14, 28].

Our deployment also helped us to identify that Crowd Coach has the potential of assisting workers to venture into new work areas. Future work could explore how peer coaching might facilitate integrating minorities into areas where they have traditionally not been represented.

In our real world deployment, we identified that workers tended to upvote coaching encouraging the use of intelligent interfaces to enhance their work. Crowd workers may not have been cognizant of existing technology that could support them in their work; and may have therefore appreciated learning about new tools to enhance their work. Alternately, it could also be that the people who are choosing to engage in crowd work have higher interests in new technologies, as previous research has also suggested [13, 54]. Nonetheless, we believe there is value in exploring intelligent interfaces that support crowd workers in developing their personal skills. For example, while a substantial portion of online microtasks focus on creating training data sets for machine learning (to give just a few examples, [12, 35, 59]), in the future we plan to explore tools for helping workers to integrate machine learning into their workflow to have better work and learning experiences [84], such as gaining a higher salary.

Our real-world deployment also uncovered there were tensions between wanting too specific or too general coaching. Some workers preferred to have coaching that was tailored for the particular HIT they were doing; while other workers preferred more general advice. In the future we will explore interfaces that can facilitate labeling coaching into either “specific” or “general” to help crowd workers better control the type of coaching they receive. We noted that workers did not appear to value having “mainstream” or “alternative” type coaching. It seemed that knowing the generalizability of the coaching they received was more important.

Similar to other crowd powered systems [41], in the future, we also plan to run a longitudinal deployment of Crowd Coach to study the dynamics that emerge with workers as they use the system long term. We do not yet know whether workers may become reluctant to coach over time if they feel they are losing opportunities in so doing. For example, if a worker coaches about how to best deal with a particular requester, a greater supply of workers may then be able to successfully complete the work of such requester, thereby potentially crowding out the original Turker. Future work could thus explore the type of coaching that workers decide to limit (not share). There might also be opportunities to innovate and design new incentives to fill the void left by possible task or requester "hoarding" by leveraging the benefits of coaching. Future work could quantify the benefits of being a crowd coach long term.

We also plan to study the system’s sustainability. It is unclear whether workers will continuously have coaching to give each other, or if there is a finite number of tips that can be given. Crowd work is continuously evolving [34]. We therefore believe that crowd workers will likely always have new advice to provide. We also believe that as workers evolve in their careers, some workers will be more likely to start assuming coaching roles and want to share their knowledge, regardless of what might exist previously. For example, Turkopticon was published in 2013 and it contains a lot of reviews of requesters [43]. Still, crowd workers continually provide new reviews of requesters to it. In the future, we are also interested in exploring how peer advice could be used to help crowd workers directly make better wages, or learn how to delegate work for which they lack the skills to complete [60].

We are also interested in the type of peer communication that workers will have with each other long term. Could there be a switch from communication focused on coaching each other to communicating focused on organizing labor unions? Prior work has shown how enabling communication between peers can lead to activism [40]. It is unclear whether systems like Crowd Coach might facilitate workers’ rights movements.

Limitations. The insights from this paper are limited by the methodology used and the population studied. While for our field experiment we used real tasks from crowd marketplaces and recruited real novice crowd workers, our skill development results might not yet generalize or apply to crowd work at large. Moreover, the pretest in our field experiment to guarantee that all the participants have similar skills means that we primarily evaluated the skill development benefits of our system with novices; we cannot speak to whether Crowd Coach could be suitable for more experienced workers, even though our ultimate goal is to provide a tool that can scale to different levels of experience. We did try however to address this issue by conducting a real world deployment where a range of hundreds of workers used our system on tens of different tasks. The deployment we conducted may also have novelty effects that need to be studied in greater depth in future work by conducting longitudinal studies.

6 CONCLUSION

This paper introduced Crowd Coach, a Chrome extension to promote workers' personal skill building by supporting peer coaching while on the job. Crowd Coach uses peer micro-advice and a twofold reputation mechanism to help workers personally improve their skills. Crowd Coach differentiates itself from prior work by helping crowd workers to improve their skills by depending on the crowd workers themselves rather than requesters or experts. Moreover, Crowd Coach innovates on related work by embedding the help mechanisms within tasks so that work growth can be achieved while on the job, and by supporting workers to provide and receive the coaching through micro-coaching snippets rather than lengthy and comprehensive assistance. In this way, our system grounds itself in the practical concerns and constraints that both crowd workers and requesters face, such as limited time and resources available to devote to improving skills or task workflow designs. The present study sets the stage for future systems that focus on creating rewarding labor experiences where crowd workers can personally grow while on the job.

Acknowledgments. This work was partially supported by FY2018 WV Research Challenge Grant award, J. Wayne and Kathy Richards Faculty Fellowship in Engineering and Research Grant from Leidos Laboratories. Special thanks AMT workers and the anonymous CSCW reviewers for their assistance in this research.

REFERENCES

[1] Ali Alkhatib, Michael S Bernstein, and Margaret Levi. 2017. Examining crowd work and gig work through the historical lens of piecework. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 4599–4616.

[2] Walter Ángel, Saiph Savage, and Nataly Moreno. 2015. Participatory Stoves: Designing Renewable Energy Technologies for the Rural Sector. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. ACM, 259–262.

[3] Thomas A Angelo and K Patricia Cross. 1993. Classroom assessment techniques: A handbookfor college teachers. *San Francisco: Jossey-Bas* (1993).

[4] Fabricio Balcazar, Bill L Hopkins, and Yolanda Suarez. 1985. A critical, objective review of performance feedback. *Journal of Organizational Behavior Management* 7, 3-4 (1985), 65–89.

[5] Martina Balestra, Coye Cheshire, Ofer Arazy, and Oded Nov. 2017. Investigating the Motivational Paths of Peer Production Newcomers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6381–6385. https://doi.org/10.1145/3025453.3026057

[6] Janine Berg. 2015. Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comparative Labor Law and Policy Journal* 37 (2015), 543.

[7] Jeffrey P Bigham, Raja Kushalnagar, Ting-Hao Kenneth Huang, Juan Pablo Flores, and Saiph Savage. 2017. On How Deaf People Might Use Speech to Control Devices. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 383–384.

[8] Jeffrey P Bigham, Kristin Williams, Nila Banerjee, and John Zimmerman. 2017. Scopist: Building a Skill Ladder into Crowd Transcription. (2017).

[9] Stephen Billett. 2001. Learning through work: workplace affordances and individual engagement. *Journal of Workplace Learning* 13, 5 (2001), 209–214.

[10] David Boud, Ruth Cohen, and Jane Sampson. 1999. Peer learning and assessment. *Assessment & Evaluation in Higher Education* 24, 4 (1999), 413–426.

[11] Daren C Brabham. 2008. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence* 14, 1 (2008), 75–90.

[12] Scott R Braithwaite, Christophe Giraud-Carrier, Josh West, Michael D Barnes, and Carl Lee Hanson. 2016. Validating machine learning algorithms for Twitter data against established measures of suicidality. *JMIR Mental Health* 3, 2 (2016).

[13] Robin Brewer, Meredith Ringel Morris, and Anne Marie Piper. 2016. Why would anybody do this?: Understanding older adults' motivations and challenges in crowd work. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2246–2257.

[14] Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann.

[15] Carrie J Cai, Shamsi T Iqbal, and Jaime Teevan. 2016. Chain reactions: The impact of order on microtask chains. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3143–3154.

[16] Chris Callison-Burch. 2014. Crowd-workers: Aggregating information across turkers to help them find higher paying work. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

[17] Patricia A Carlson and Frederick C Berry. 2003. Calibrated peer review/sup TM/and assessing learning outcomes. In *Frontiers in Education Conference*. IEEE, F3E1–6.

[18] Susan Cartwright and Nicola Holmes. 2006. The meaning of work: The challenge of regaining employee engagement and reducing cynicism. *Human Resource Management Review* 16, 2 (2006), 199–208.

[19] Joel Chan, Steven Dang, and Steven P Dow. 2016. Comparing different sensemaking approaches for large-scale ideation. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 2717–2728.

[20] Jesse Chandler, Gabriele Paolacci, and Pam Mueller. 2013. *Risks and Rewards of Crowdsourcing Marketplaces*. Springer New York, New York, NY, 377–392. https://doi.org/10.1007/978-1-4614-8806-4_30

[21] Chun-Wei Chiang, Eber Betanzos, and Saiph Savage. 2018. Exploring Blockchain for Trustful Collaborations between Immigrants and Governments. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, LBW531.

[22] Derrick Coetzee, Seongtaek Lim, Armando Fox, Bjorn Hartmann, and Marti A Hearst. 2015. Structuring interactions for large-scale synchronous peer learning. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1139–1152.

[23] Xuefei Nancy Deng and KD Joshi. 2013. Is Crowdsourcing a Source of Worker Empowerment or Exploitation? Understanding Crowd Workers' Perceptions of Crowdsourcing Career. (2013).

[24] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G Ipeirotis, and Philippe Cudré-Mauroux. 2015. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 238–247.

[25] Martin Dittus and Licia Capra. 2017. Private Peer Feedback As Engagement Driver in Humanitarian Mapping. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 40 (Dec. 2017), 18 pages. https://doi.org/10.1145/3134675

[26] Mira Dontcheva, Robert R Morris, Joel R Brandt, and Elizabeth M Gerber. 2014. Combining crowdsourcing and learning to improve engagement and performance. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3379–3388.

[27] Shayan Doroudi, Ece Kamar, Emma Brunskill, and Eric Horvitz. 2016. Toward a learning science for complex crowdsourcing tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2623–2634.

[28] Steven P Dow, Kate Heddleston, and Scott R Klemmer. 2009. The efficacy of prototyping under time constraints. In *Proceedings of the seventh ACM conference on Creativity and cognition*. ACM, 165–174.

[29] David Durward, Ivo Blohm, and Jan Marco Leimeister. 2016. Crowd work. *Business & Information Systems Engineering* 58, 4 (2016), 281–286.

[30] Roberta L Duyff. 1999. The value of lifelong learning: key element in professional career development. *Journal of the Academy of Nutrition and Dietetics* 99, 5 (1999), 538.

[31] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A Landay. 2011. MicroMandarin: mobile language learning in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3169–3178.

[32] Nancy Falchikov and Judy Goldfinch. 2000. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of educational research* 70, 3 (2000), 287–322.

[33] Benjamin V Hanrahan, Jutta K Willamowski, Saiganesh Swaminathan, and David B Martin. 2015. TurkBench: Rendering the market for Turkers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1613–1616.

[34] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham. 2018. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on*

Human Factors in Computing Systems. ACM, 449.
[35] Brent Harrison and Mark O Riedl. 2016. Learning from stories: using crowdsourced narratives to train virtual agents. *Proceedings of the Artificial Intelligence and Interactive Digital Entertainment* (2016), 183–189.
[36] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of Educational Research* 77, 1 (2007), 81–112.
[37] Yurong He, Jennifer Preece, Carol Boston, Anne Bowser, Derek Hansen, and Jen Hammock. 2014. The Effects of Individualized Feedback on College Students' Contributions to Citizen Science. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW Companion '14)*. ACM, New York, NY, USA, 165–168. https://doi.org/10.1145/2556420.2556484
[38] Pamela J Hinds, Michael Patterson, and Jeffrey Pfeffer. 2001. Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of applied psychology* 86, 6 (2001), 1232.
[39] John B. Horrigan. 2016. Civic engagement in the digital age. *Pew Internet & American Life Project* (2016).
[40] Philip N Howard, Saiph Savage, Claudia Flores Saviaga, Carlos Toxtli, and Andrés Monroy-Hernández. 2016. Social media, civic engagement, and the slacktivism hypothesis: Lessons from Mexico's "El Bronco". *Journal of International Affairs* 70, 1 (2016).
[41] Ting-Hao Kenneth Huang, Walter S Lasecki, Amos Azaria, and Jeffrey P Bigham. 2016. " Is There Anything Else I Can Help You With?" Challenges in Deploying an On-Demand Crowd-Powered Conversational Agent. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*.
[42] Kazushi Ikeda and Michael S Bernstein. 2016. Pay It Backward: Per-Task Payments on Crowdsourcing Platforms Reduce Productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4111–4121.
[43] Lilly C Irani and M Silberman. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 611–620.
[44] Lilly C Irani and M Silberman. 2016. Stories we tell about labor: Turkopticon and the trouble with design. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 4573–4586.
[45] Nila Banerjee Jeffrey P. Bigham, Kristin Williams and John Zimmerman. 2017. Scopist: Building a Skill Ladder into Crowd Work. In *Proceedings of the Web for All Conference (W4A '17)*. ACM, New York, NY, USA, 10.
[46] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money. Worker Motivation in Crowdsourcing-A Study on Mechanical Turk.. In *AMCIS*, Vol. 11. 1–11.
[47] Clare Kelliher and Deirdre Anderson. 2008. For better or for worse? An analysis of how flexible working practices influence employees' perceptions of job quality. *The International Journal of Human Resource Management,* 19, 3 (2008), 419–431.
[48] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1301–1318.
[49] Avraham N Kluger and Angelo DeNisi. 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin* 119, 2 (1996), 254.
[50] Kenneth R Koedinger, Albert T Corbett, and Charles Perfetti. 2012. The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive science* 36, 5 (2012), 757–798.
[51] Anand Kulkarni, Matthew Can, and Björn Hartmann. 2012. Collaboratively Crowdsourcing Workflows with Turkomatic. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 1003–1012. https://doi.org/10.1145/2145204.2145354
[52] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R Klemmer. 2013. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 6 (2013), 33.
[53] Chinmay E Kulkarni, Michael S Bernstein, and Scott R Klemmer. 2015. PeerStudio: rapid peer feedback emphasizes revision and improves performance. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 75–84.
[54] Airi Lampinen, Victoria Bellotti, Coye Cheshire, and Mary Gray. 2016. CSCW and theSharing Economy: The Future of Platforms as Sites of Work Collaboration and Trust. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. ACM, 491–497.
[55] Gary P Latham and Edwin A Locke. 1991. Self-regulation through goal setting. *Organizational behavior and human decision processes* 50, 2 (1991), 212–247.
[56] Carol A Lundberg. 2003. The influence of time-limitations, faculty, and peer relationships on adult student learning: A causal model. *The Journal of Higher Education* 74, 6 (2003), 665–688.
[57] David Martin, Benjamin V Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 224–235.
[58] Winter Mason and Duncan J Watts. 2010. Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter* 11, 2 (2010), 100–108.

[59] Alexander H Miller, Will Feng, Adam Fisch, Jiasen Lu, Dhruv Batra, Antoine Bordes, Devi Parikh, and Jason Weston. 2017. Parlai: A dialog research software platform. *arXiv preprint arXiv:1705.06476* (2017).

[60] Meredith Ringel Morris, Jeffrey P Bigham, Robin Brewer, Jonathan Bragg, Anand Kulkarni, Jessie Li, and Saiph Savage. 2017. Subcontracting microwork. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1867–1876.

[61] Melissa M Nelson and Christian D Schunn. 2009. The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science* 37, 4 (2009), 375–401.

[62] Jeffrey Nichols and Jeon-Hyung Kang. 2012. Asking questions of targeted strangers on social networks. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 999–1002.

[63] Jeffrey Nichols, Michelle Zhou, Huahai Yang, Jeon-Hyung Kang, and Xiao Hua Sun. 2013. Analyzing the quality of information solicited from targeted strangers on social media. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 967–976.

[64] Jon Noronha, Eric Hysen, Haoqi Zhang, and Krzysztof Z Gajos. 2011. Platemate: crowdsourcing nutritional analysis from food photographs. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–12.

[65] Scott Novotney and Chris Callison-Burch. 2010. Cheap, fast and good enough: Automatic speech recognition with non-expert transcription. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 207–215.

[66] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. (2010).

[67] Melissa M Patchan and Christian D Schunn. 2015. Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality. *Instructional Science* 43, 5 (2015), 591–614.

[68] Sunil Ramlall. 2004. A review of employee motivation theories and their implications for employee retention within organizations. *Journal of American Academy of Business* 5, 1/2 (2004), 52–63.

[69] Alex Rosenblat and Luke Stark. 2016. Algorithmic labor and information asymmetries: A case study of Uber's drivers. (2016).

[70] Joel Ross, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. 2009. Who are the turkers? Worker demographics in amazon mechanical turk. *Department of Informatics, University of California, Irvine, USA, Tech. Rep* (2009).

[71] 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.

[72] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. 2016. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 813–822.

[73] Jodi Schneider, Bluma S. Gelley, and Aaron Halfaker. 2014. Accept, Decline, Postpone: How Newcomer Productivity is Reduced in English Wikipedia by Pre-publication Review. In *Proceedings of The International Symposium on Open Collaboration (OpenSym '14)*. ACM, New York, NY, USA, Article 26, 10 pages. https://doi.org/10.1145/2641580.2641614

[74] M Silberman, Lilly Irani, and Joel Ross. 2010. Ethics and tactics of professional crowdwork. *XRDS: Crossroads, The ACM Magazine for Students* 17, 2 (2010), 39–43.

[75] M. Six Silberman, Joel Ross, Lilly Irani, and Bill Tomlinson. 2010. Sellers' Problems in Human Computation Markets. In *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10)*. ACM, New York, NY, USA, 18–21. https://doi.org/10.1145/1837885.1837891

[76] Ryo Suzuki, Niloufar Salehi, Michelle S Lam, Juan C Marroquin, and Michael S Bernstein. 2016. Atelier: Repurposing expert crowdsourcing tasks as micro-internships. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2645–2656.

[77] William Thies, Aishwarya Ratan, and James Davis. 2011. Paid crowdsourcing as a vehicle for global development. In *CHI Workshop on Crowdsourcing and Human Computation*.

[78] David Tinapple, Loren Olson, and John Sadauskas. 2013. CritViz: Web-based software supporting peer critique in large creative classrooms. *Bulletin of the IEEE Technical Committee on Learning Technology* 15, 1 (2013), 29.

[79] Bonnie Urciuoli. 2008. Skills and selves in the new workplace. *American Ethnologist* 35, 2 (2008), 211–228.

[80] Rajan Vaish, Keith Wyngarden, Jingshu Chen, Brandon Cheung, and Michael S Bernstein. 2014. Twitch crowdsourcing: crowd contributions in short bursts of time. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 3645–3654.

[81] Niels van Doorn. 2017. Platform labor: on the gendered and racialized exploitation of low-income service work in the 'on-demand' economy. *Information, Communication & Society* 20, 6 (2017), 898–914.

[82] 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 (CSCW '15)*. ACM, New York, NY, USA, 405–416. https://doi.org/10.1145/2675133.2675219

[83] Mark E Whiting, Dilrukshi Gamage, Snehalkumar S Gaikwad, Aaron Gilbee, Shirish Goyal, Alipta Ballav, Dinesh Majeti, Nalin Chhibber, Angela Richmond-Fuller, Freddie Vargus, et al. 2016. Crowd guilds: Worker-led reputation and feedback on crowdsourcing platforms. *arXiv preprint arXiv:1611.01572* (2016).

[84] 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.

[85] Bill Williamson. 1998. Lifeworlds and Learning. Essays in the Theory, Philosophy and Practice of Lifelong Learning. (1998).

[86] Yu-Chun Grace Yen, Steven P Dow, Elizabeth Gerber, and Brian P Bailey. 2016. Social Network, Web Forum, or Task Market?: Comparing Different Crowd Genres for Design Feedback Exchange. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, 773–784.

[87] Dongwook Yoon, Nicholas Chen, Bernie Randles, Amy Cheatle, Corinna E. Löckenhoff, Steven J. Jackson, Abigail Sellen, and François Guimbretière. 2016. RichReview++: Deployment of a Collaborative Multi-modal Annotation System for Instructor Feedback and Peer Discussion. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 195–205. https://doi.org/10.1145/2818048.2819951

[88] Haiyi Zhu, Steven P Dow, Robert E Kraut, and Aniket Kittur. 2014. Reviewing versus doing: Learning and performance in crowd assessment. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 1445–1455.

[89] Haiyi Zhu, Amy Zhang, Jiping He, Robert E. Kraut, and Aniket Kittur. 2013. Effects of Peer Feedback on Contribution: A Field Experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2253–2262. https://doi.org/10.1145/2470654.2481311

Received April 2018; revised July 2018; accepted September 2018