Exploring the Effects of Goal Setting When Training for Complex Crowdsourcing Tasks (Extended Abstract)

Amy Rechkemmer , **Ming Yin**

Purdue University

{arechke, mingyin}@purdue.edu

Abstract

Training is one way of enabling novice workers to work on complex crowdsourcing tasks. Based on goal setting theory in psychology, we conduct a randomized experiment to study whether and how setting different goals—including performance goal, learning goal, and behavioral goal—when training workers for a complex crowdsourcing task affects workers' learning perception, learning gain, and post-training performance. We find that setting different goals during training significantly affects workers' learning perception, but does not have an effect on learning gain or post-training performance. Further, exploratory analysis helps shed light on when and why various goals may or may not work in the crowdsourcing context.

1 Introduction

Online crowdsourcing has made it easy for researchers and professionals alike to collect human-specific knowledge quickly and easily. Although early practice of crowdsourcing often solicited human labor on tasks that require only basic human skills, there have been substantial efforts recently to accommodate complex tasks requiring sophisticated domain knowledge in crowdsourcing settings.

A variety of approaches have been developed to enable the completion of complex tasks by the crowd. For example, workers are trained within a complex task to be equipped with necessary knowledge and strategies for completing the task. A wide range of training methods have been studied, including training by examples and gold standard tasks [Le *et al.*, 2010; Mitra *et al.*, 2015; Liu *et al.*, 2016], training by self-evaluation, expert assessment or peer feedback [Dow *et al.*, 2012; Zhu *et al.*, 2014; Doroudi *et al.*, 2016], and training by communication with mentors and peers [Chen *et al.*, 2019; Tang *et al.*, 2019; Suzuki *et al.*, 2016].

In traditional educational or organizational settings, an important aspect in motivating people to learn is to *set goals*. Multiple types of goals have been studied, including performance goals (i.e., goals that specify the targeted end results of learning), learning goals (i.e., goals that state the range of knowledge the learner aims to master), and behavioral goals (i.e., goals that describe a set of desirable be-

haviors the learner should follow). The effects of goals are found to be dependent on both the type of goal and who sets the goal [Latham and Brown, 2006; Seijts *et al.*, 2004; Clark *et al.*, 2016]. Moreover, researchers have identified that different people have a different "goal orientation," that is, the primary factors that motivate the individual [Dweck, 1986; Bell and Kozlowski, 2002], and that individuals with different types of goal orientation also respond to various types of goals differently [Seijts *et al.*, 2004; Button *et al.*, 1996].

The effects of setting goals when training crowd workers for complex tasks, however, are under-explored. To fill in this gap, we design and conduct an experiment on Amazon Mechanical Turk that trains crowd workers to perform a task requiring substantial domain knowledge. We consider 7 experimental treatments in our study that differ in whether and what type of goal is set during the training stage, as well as who sets the goal. To the best of our knowledge, this is the first study that investigates the motivating effects of goals on training workers towards better learning and performance in complex crowdsourcing tasks.

Overall, we find that setting different goals when training crowd workers in a complex task significantly influences workers' learning perceptions, but does not affect their learning gain or post-training task performance. However, for the subset of workers who have high learning goal orientation, setting learning goals for them does lead to higher learning gain. Additionally, workers given a challenging behavioral goal perform significantly more of the desirable behavior, and workers who performed more of this behavior had higher learning gain and post-training performance, but only when they did so of their own volition. We conclude by discussing the implications of our findings.

2 Study Design

To understand the effects of setting goals in the training stage of complex crowdsourcing tasks, we designed and conducted an experiment on Amazon Mechanical Turk (MTurk). Our main research questions were:

- **Q1**: How does setting different goals affect workers' learning perceptions during training?

- **Q2**: How does setting different goals affect workers' learning gain during training and performance on tasks after training?
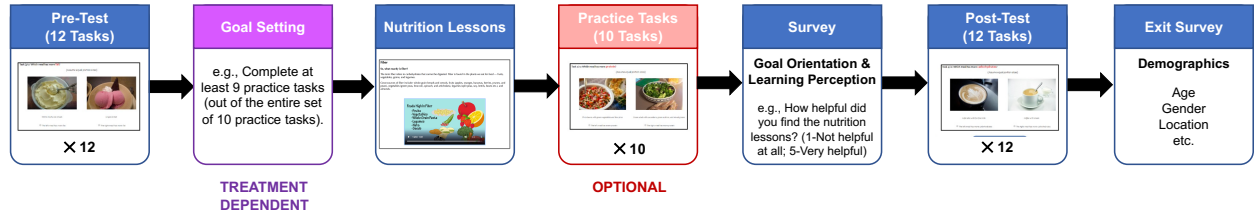
Figure 1: A diagram of sections that workers went through in phase 1 of our experiment.

2.1 Experimental Tasks

We used a task that asked workers to identify nutritional components in meals. In each task, a worker was given two photographs of meals along with descriptions of the main ingredients in each meal. The worker was asked to identify which of the two meals contained more of a specified nutritional component. Four nutritional components—fat, fiber, protein, and carbohydrates—were examined in these tasks. Photographs used were taken from Burgermaster *et al.* [2017]. We chose this task for our study as it requires substantial nutritional knowledge, which is not a common skill among laypeople.

2.2 Experimental Procedure

Our experiment was divided into three phases. Phase 1 was the "training" phase and was used to prepare workers with necessary nutritional knowledge before they complete actual nutrition tasks. Phases 2 and 3 were the work phases in which workers could use the knowledge they learned from phase 1 to complete a sequence of nutrition tasks. Separate pools of tasks were created to be sampled from for each of the phases.

Phase 1

Figure 1 displays the overall flow of phase 1. A worker started by completing a session of 12 randomly-sampled nutrition tasks, which we refer to as the *pre-test*. Upon completion of these tasks, we told workers their accuracy in these 12 tasks. Next, prior to taking the nutrition lessons, the worker might be given a goal or be asked to set a goal for herself for the nutrition lessons depending on the treatment the worker was assigned (see more details in Section 2.3). The worker was then required to go through a nutrition lesson for each nutritional component. After the nutrition lessons, the worker could take up to 10 practice tasks, which were in the same form as the nutrition tasks. For each practice task that the worker took, we provided her feedback on both the correct answer and an explanation.

At the end of phase 1, the worker first took a questionnaire on her goal orientation. We adopted the scales from Button *et al.* [1996] to measure the worker's goal orientation in terms of performance (i.e., the level of motivation towards achieving high performance) or learning (i.e., the level of motivation towards learning new things). Second, she answered two survey questions on a 5-point scale regarding her perception of learning in the nutrition lessons:

- *Helpfulness*: How helpful did you find the nutrition lessons?
- *Learning*: How much do you feel you have learned from the nutrition lessons you went through earlier?

Finally, the worker completed another session of 12 randomly-sampled *post-test* nutrition tasks, and filled in an exit survey on their demographics.

Phases 2 and 3

Phase 2 happened two days after phase 1, and phase 3 happened one week after phase 1. Regardless of the worker's treatment, in both phases 2 and 3, she was asked to complete a random sequence of 12 previously unseen nutrition tasks. We did not provide any feedback on answer accuracy for tasks in phases 2 and 3. Thus, the worker's accuracy in phases 2 and 3 reflected her performance on real-world tasks, either shortly after the training or a while after receiving the training.

2.3 Experimental Treatments

We considered a 2×3 design along two factors: the *initiator* of the goal and the *type* of the goal. The initiator of the goal can be either the worker herself or the requester. When the initiator of the goal was the worker, she was asked to set a goal for herself, but when the initiator of the goal was the requester, the worker was given a goal that was pre-determined by us. We considered three types of goals in our experiment:

- *Performance goal*: This goal specifies the number of *post-test* tasks that the worker should answer *correctly*. When the requester set the goal, it was to answer at least 10 out of 12 possible post-test questions correctly. When the worker set the goal, she could choose any integer between 0 and 12 of post-test tasks to answer correctly.

- *Learning goal*: This goal specifies the kind of knowledge the worker aims to learn from the nutrition lessons. When the requester set the goal, it was stated as "Learn and recognize the types of foods that are high in carbohydrates, protein, fiber, and fat." Workers asked to set their own goal used free-form language to create goals.

- *Practice goal*: This goal is an operationalization of the "behavioral goal" and specifies the number of practice tasks that the worker should complete after taking the nutrition lessons. When the requester set the goal, it was to complete at least 9 out of 10 practice tasks. When the worker set the goal, she could choose any integer between 0 and 10 of practice tasks to complete.

Finally, we also included a control treatment where the worker was not given a goal nor asked to set a goal for herself. Together, we had 7 treatments in this experiment. Our study was open only to U.S. workers who had completed at least 500 HITs on MTurk previously. 659 workers participated in phase 1, and they were *randomly* assigned to one
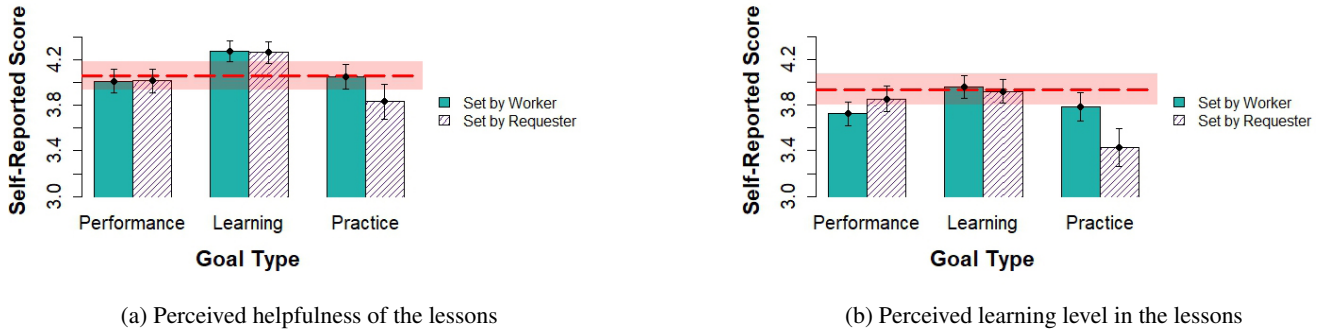
(a) Perceived helpfulness of the lessons

(b) Perceived learning level in the lessons

Figure 2: Workers' perceptions of learning across different treatments. The mean value of self-reported scores for each treatment is plotted, and error bars represent the standard errors of the mean. For the control treatment, the dashed horizontal lines represent the mean values, and the standard errors of the mean are shown by the red shaded areas.

of the treatments upon arrival. Among all phase 1 participants, 581 workers further took the phase 2 experiment and 558 workers took the phase 3 experiment.

3 Results

We first analyze the data to answer our main research questions. Then, we conduct additional exploratory analyses to understand why and under which conditions certain goals are effective or not effective.

3.1 Q1: The Impact on Learning Perception

We measure workers' learning perceptions using their self-reported scores on the helpfulness of the training material and the amount they have learned in the nutrition lessons, and Figures 2a and 2b show the comparison on these two metrics, respectively. A one-way ANOVA test across all 7 treatments suggests a marginally significant difference in workers' scores on the helpfulness of nutrition lessons ($p = 0.096$) and workers' perceived levels of learning ($p = 0.075$). Thus, setting different goals does affect workers' learning perceptions.

To further understand how the initiator and type of the goal affect workers' perceptions of learning, we conducted a two-way ANOVA test on the data obtained from all but the control treatment. Doing so, we found that the type of the goal has a significant effect on both the reported helpfulness of the lessons ($p = 0.008$) and the perceived level of learning ($p = 0.036$). Post-hoc Tukey HSD test results suggest that workers with a learning goal reported the lessons as significantly more helpful than workers with other goal types (performance: $p = 0.037$, practice: $p = 0.015$), and they also perceived themselves as having learned significantly more from the lessons compared to workers with a practice goal ($p = 0.028$). However, the initiator of the goal does not have a significant effect on either metric.

3.2 Q2: The Impact on Learning Gain and Post-Training Performance

We define a worker's learning gain as the difference in the number of pre-test and post-test questions that she answered correctly in phase 1. A one-way ANOVA test across all 7 treatments suggests no statistically significant difference in

learning gain across different treatments, and when we conducted a two-way ANOVA test on the 6 non-control treatments, we still found that neither the initiator of the goal nor the type of the goal has any significant impact on how much the worker actually learned. We then performed the same analysis on our data collected from phases 2 and 3 to understand the impact of setting goals on post-training performance. Again, we found no significant effects of goals.

Overall, setting different goals when training our workers for the nutrition task does *not* lead to significantly different learning outcomes or post-training task performance.

3.3 Exploratory Analysis

In contrast with findings on the effects of goals in prior literature, we did not find that setting goals during training had any effect on learning gain or post-training performance. This motivated us to look deeper into when and why various goals may work or not work in the crowdsourcing context.

On the one hand, previous research suggests that individuals have different types of goal orientation [Dweck, 1986; Bell and Kozlowski, 2002] and may respond to various goals differently depending on whether or not the goal *matches* with their goal orientation. On the other hand, goals like the behavioral goal are designed to encourage the adoption of desirable behavior. Since we did not see that setting behavioral goals for workers leads to any significant improvement in learning gain or performance, we seek to explore the reason why here. This led us to the following additional questions:

- Does setting a goal for workers that matches with their goal orientation lead to higher levels of learning gain and post-training performance?

- Does setting a practice goal lead to more practice tasks being completed, and does completing more practice tasks associate with higher levels of learning gain and post-training performance?

The Role of Goal Orientation on Goal Effectiveness

We used a median split to classify each worker as "high" or "low" on performance (or learning) goal orientation based on her responses to the goal orientation scales during phase 1. Figure 3a displays how setting different goals in the training stage affects the learning gain of workers with high performance goal orientation. We found that setting a performance
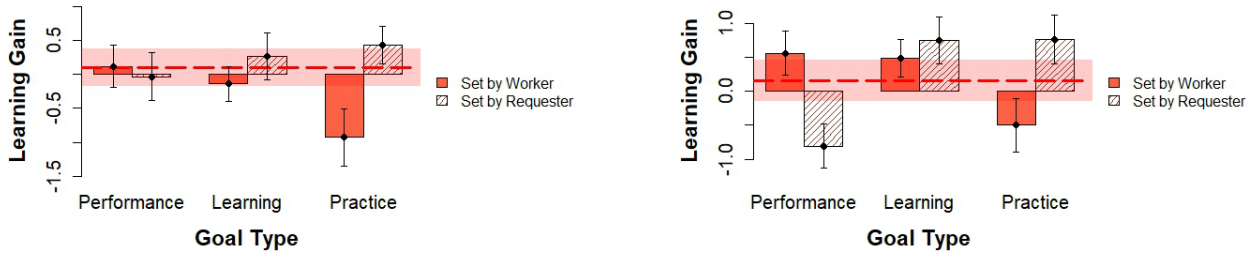
(a) Workers with high performance goal orientation

(b) Workers with high learning goal orientation

Figure 3: The learning gain across different treatments for workers with high performance/learning goal orientation. The mean value of the learning gain for each treatment is plotted, and error bars represent the standard errors of the mean. Dashed lines and red shaded areas show the mean values and standard errors of the mean for learning gain in the control treatment.

goal for workers who have high performance goal orientation does *not* lead to significantly higher levels of learning gain or post-training performance compared to the cases when no goal or other types of goals are set.

On the other hand, Figure 3b shows the impact of goals on learning gain when restricted to workers who are more motivated to learn new things. Results of a one-way ANOVA test suggest a statistically significant difference in learning gain across all 7 treatments ($p = 0.004$), and a two-way ANOVA test suggests that the type of goal has a significant effect ($p = 0.049$). Post-hoc Tukey HSD tests revealed that for workers who are highly motivated to learn new things, a learning goal leads to higher levels of learning gain compared to a performance goal when the goal is set by the requester ($p = 0.025$). The higher levels of learning gain, however, do not translate into higher post-training performance.

In sum, we found that matching a performance goal to workers who are highly oriented for performance does not improve either the learning gain or post-training performance, while matching a learning goal to workers who are highly oriented for learning improves the learning gain, but does not affect the post-training performance.

Why Behavioral Goals Don't Work?

We now move on to explore why behavioral goals seem to be ineffective in complex crowdsourcing task training. First, we note that when workers were asked to set a practice goal for themselves, they tended to set an "easy" goal—on average, they aimed at completing 4.3 practice tasks, while workers who were given a practice goal were told to try to complete at least 9 practice tasks. We observed a statistically significant difference ($p < 0.001$) in the number of practice tasks that workers of different treatments completed, and in particular, post-hoc Tukey HSD tests show that workers who were given a practice goal completed significantly more practice tasks than workers in all other treatments ($p < 0.005$).

We then ask whether completing more practice tasks associates with higher learning gain and better post-training performance for workers. Conducting two-sample t-tests, we confirmed that workers who completed more practice tasks not only learned more during the training stage, but also achieved better performance in both phase 2 and phase 3

($p < 0.001$ for all three comparisons).

This result is puzzling given that we did not find the practice goal had any effect on learning gain or post-training performance. An in-depth analysis of the data suggests one possible explanation for this discrepancy—while workers who were given a challenging practice goal did complete more practice tasks, the increase in learning gain (or post-training performance) they obtained from the extra practice is *less* than the increase that workers from other conditions experienced. We conjecture that many of our workers assigned a challenging practice goal chose to satisfice, leading to them meeting the goal even if they did not reap the benefits that those who completed the practice of their own volition did.

4 Conclusion

Our study shows that setting different goals when training workers for complex crowdsourcing tasks indeed changes worker's learning perceptions, but it has no significant impact on workers' learning gain or post-training performance. Through exploratory analysis, we find that certain types of goals can motivate workers to improve learning gain and post-training performance, but only under certain conditions. We conjecture that the difference in our results from those of prior literature may be due to crucial differences between the traditional educational and organizational environments of goal setting literature and the crowdsourcing environment.

One of the lessons that we learned through this study is that when training workers for complex crowdsourcing tasks, the right type of goal needs to be set for the right kind of workers in the right way, suggesting the potential for *personalizing* goals. Another lesson is the need to improve the design of behavioral goals in crowdsourcing settings, as the practice goals in our experiment were not very effective in influencing learning and performance, but we were able to see the potential of adopting desirable behavior.

A direction of future work is to explore the design space of goals and examine their effectiveness along other key design dimensions, such as the connection between goal attainment with different incentives. Exploring the use of goal setting on different online crowdsourcing platforms that have different inherent incentives (e.g., citizen science platforms) is an additional direction to pursue.

References

[Bell and Kozlowski, 2002] Bradford S Bell and WJ Kozlowski. Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology*, 87(3):497, 2002.

[Burgermaster *et al.*, 2017] Marissa Burgermaster, Krzysztof Z Gajos, Patricia Davidson, and Lena Mamykina. The role of explanations in casual observational learning about nutrition. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4097–4145. ACM, 2017.

[Button *et al.*, 1996] Scott B Button, John E Mathieu, and Dennis M Zajac. Goal orientation in organizational research: A conceptual and empirical foundation. *Organizational behavior and human decision processes*, 67(1):26–48, 1996.

[Chen *et al.*, 2019] Quanze Chen, Jonathan Bragg, Lydia B Chilton, and Dan S Weld. Cicero: Multi-turn, contextual argumentation for accurate crowdsourcing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[Clark *et al.*, 2016] Damon Clark, David Gill, Victoria Prowse, and Mark Rush. Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, pages 1–45, 2016.

[Doroudi *et al.*, 2016] Shayan Doroudi, Ece Kamar, Emma Brunskill, and Eric Horvitz. Toward a learning science for complex crowdsourcing tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2623–2634, New York, NY, USA, 2016. ACM.

[Dow *et al.*, 2012] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 1013–1022, New York, NY, USA, 2012. ACM.

[Dweck, 1986] Carol S Dweck. Motivational processes affecting learning. *American psychologist*, 41(10):1040, 1986.

[Latham and Brown, 2006] Gary P Latham and Travor C Brown. The effect of learning vs. outcome goals on self-efficacy, satisfaction and performance in an mba program. *Applied Psychology*, 55(4):606–623, 2006.

[Le *et al.*, 2010] John Le, Andy Edmonds, Vaughn Hester, and Lukas Biewald. Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. In *In SIGIR 2010 workshop*, pages 21–26, 2010.

[Liu *et al.*, 2016] Angli Liu, Stephen Soderland, Jonathan Bragg, Christopher H. Lin, Xiao Ling, and Daniel S. Weld. Effective crowd annotation for relation extraction. In *HLT-NAACL*, 2016.

[Mitra *et al.*, 2015] Tanushree Mitra, C.J. Hutto, and Eric Gilbert. Comparing person- and process-centric strategies for obtaining quality data on amazon mechanical turk. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1345–1354, New York, NY, USA, 2015. ACM.

[Seijts *et al.*, 2004] Gerard H Seijts, Gary P Latham, Kevin Tasa, and Brandon W Latham. Goal setting and goal orientation: An integration of two different yet related literatures. *Academy of management journal*, 47(2):227–239, 2004.

[Suzuki *et al.*, 2016] Ryo Suzuki, Niloufar Salehi, Michelle S Lam, Juan C Marroquin, and Michael S Bernstein. Atelier: Repurposing expert crowdsourcing tasks as micro-internships. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2645–2656. ACM, 2016.

[Tang *et al.*, 2019] Wei Tang, Ming Yin, and Chien-Ju Ho. Leveraging peer communication to enhance crowdsourcing. In *The World Wide Web Conference*, WWW '19, pages 1794–1805, New York, NY, USA, 2019. ACM.

[Zhu *et al.*, 2014] Haiyi Zhu, Steven P Dow, Robert E Kraut, and Aniket Kittur. Reviewing versus doing: learning and performance in crowd assessment. *Conference on Computer Supported Cooperative Work*, pages 1445–1455, 2014.