International Journal on Cybernetics & Informatics (IJCI) Vol.15, No.1, February 2026

Hierarchical Worker Evaluation Based on Requester's Subjective Criteria in Open-Ended Crowdsourcing Tasks

Ryuya Itano[1], Honoka Tanitsu[1], Motoki Bamba[1], Ryota Noseyama[2], Akihito Kohiga[2], and Takahiro Koita[1]

[1] Graduate School of Science and Engineering, Doshisha University, Kyoto, Japan.

[2] Faculty of Science and Engineering, Doshisha University, Kyoto, Japan.

Abstract. Crowdsourcing assumes a transient relationship between task requesters and workers, which makes it hard for workers to improve their skills. In addition, with the emergence of AI, crowd work is shifting from simple tasks to more complex and open-ended ones, highlighting the importance of training workers to handle such tasks. Although various methods have been proposed to train and evaluate workers, a method to evaluate them in open-ended tasks among workers has not yet been established. In this study, we propose applying a hierarchical inter-worker evaluation structure based on workers' skill levels to the evaluation of open-ended tasks, and examine how closely it aligns with the requesters' subjective evaluation criteria. The experimental results showed that evaluations from workers were highly aligned with those from requesters in terms of relative worker rankings. However, the alignment was weaker for absolute scores, due to workers' tendency toward generous scoring. These findings are expected to be utilized in future research to enhance worker engagement and retention rates.

Keywords: Crowdsourcing, Worker training, Worker evaluation, Amazon Mechanical Turk

1 Introduction

Crowdsourcing is a work model that outsources various tasks to a decentralized and unspecified large group of people (crowd workers) in exchange for rewards such as money. Crowdsourcing has been widely adopted because of the following advantages: For task requesters, crowdsourcing allows them to outsource a large volume of tasks at a low cost. For workers, it allows them to work freely without time or location restrictions.

Although these advantages have promoted the development of the crowdsourcing model, it faces inherent challenges. In a crowdsourcing model, the relationship between requesters and workers is often transient. This non-persistent working relationship may slow down workers' skill development and foster low worker engagement. Low engagement discourages workers from producing consistent high-quality outputs or investing in skill improvement, which poses a threat to the long-term reliability and sustainability of the crowdsourcing model. Indeed, in order to collect consistent high-quality outputs from workers, various quality control methods have been proposed[1]. Some of these methods assume that workers are non-persistent. For example, methods using redundancy or majority voting have been developed to ensure high-quality outputs, even when workers are frequently replaced. These methods work well when the unique true answer to a task is defined and a ground truth exists. However, as tasks become increasingly complex and open-ended, and thus no longer solvable by just any worker, relying on worker redundancy will fail to secure the skilled workers capable of solving such tasks. Therefore, future quality control methods

Bibhu Dash et al: EDU, EDUTEC, CCITT - 2026
pp. 01-12, 2026. IJCI – 2026
DOI:10.5121/ijci.2026.150101

should focus on building persistent working relationships that support workers' individual skill development.

Previously, most tasks required no advanced skills and had unique true answers. However, with the emergence of generative AI models such as large language models (LLMs) and multimodal models, AI has become capable of solving tasks with unique true answers. Consequently, the importance of tasks with non-unique true answers has increased, emphasizing the need for workers to perform more complex and open-ended tasks[2]. Therefore, requesters need to train workers to ensure a stable supply of skilled workers for solving complex and open-ended tasks.

In open-ended tasks, where numerous valid answers are possible, workers are expected to produce outputs that align with the requesters' expectations. Therefore, training workers to perform well in open-ended tasks requires that requesters evaluate workers' outputs and provide appropriate feedback. Furthermore, a recent study points out that human feedback remains important for capturing requesters' expectations, even with the advancement of generative AI[3]. However, providing direct feedback to workers places a heavy workload on requesters. To address this issue, previous studies have proposed various inter-worker evaluation approaches to evaluate workers and provide them with feedback, such as peer review models and guild-like models with a hierarchical structure[4][5]. In particular, hierarchical structures are effective not only for achieving accurate inter-worker evaluation but also for enhancing engagement and commitment in real-world settings[6]. Nevertheless, these hierarchical inter-worker evaluation approaches have not been empirically validated regarding workers' ability to capture requesters' expectations (subjective evaluation criteria) in open-ended tasks. Capturing requesters' subjective evaluation criteria is essential for understanding the tacit knowledge embedded in the evaluation and for enabling workers to produce outputs that align with the requesters' intent. In turn, the requesters and workers share a common understanding, enabling the crowdsourcing to serve as a more reliable and sustainable work model for complex and open-ended tasks.

In this study, we focus on worker evaluation as a preliminary stage of worker training. We also propose applying a hierarchical inter-worker evaluation structure—which has been shown to be effective in previous studies on tasks with unique true answers—as a way to capture requesters' subjective evaluation criteria in open-ended tasks. The main contributions of this study are as follows:

- We propose applying a hierarchical inter-worker evaluation structure where workers are classified by skill levels for evaluating their outputs of open-ended tasks.
- We empirically examine how closely a hierarchical inter-worker evaluation aligns with the requesters' subjective evaluation.
- We analyze the experimental results and discuss what the hierarchical inter-worker evaluation structure can achieve and the challenges that remain for future work.

2 Related Work

Studies in crowdsourcing have explored various approaches to evaluate workers and provide them with feedback. While these methods have shown promise in improving worker evaluation and supporting skill development, their effectiveness for complex and open-ended tasks remains largely unverified.

2.1 Feedback from Requesters

Feedback directly provided by requesters has been shown to improve the quality of workers' subsequent outputs. Dow et al. developed a system that enables requesters to provide

feedback to workers and demonstrated that such feedback improves output quality [7]. Similarly, Bhattacharya et al. showed that corrective feedback from requesters not only improves work quality but also helps lower the entry barrier for new workers, thereby facilitating their skill development [8]. These studies highlight the importance of direct requester feedback as a driver of worker growth; however, such mechanisms impose additional workload on requesters, making them difficult to scale.

2.2 Peer Feedback among Workers

Peer feedback offers a scalable way to support worker evaluation and skill development without direct requester involvement. Several studies have shown that peer feedback among workers contributes to skill improvement and higher-quality outputs[9][10]. Furthermore, other studies in education and crowdsourcing have shown that not only receiving but also providing feedback contribute to workers' skill improvement[11][12]. These findings provide evidence that peer review can improve workers' skills. However, little empirical evidence exists on whether workers can accurately evaluate the skills of others in crowdsourcing settings. In the field of education, De Alfaro demonstrated that peer assessment among students can yield evaluations closely aligned with those of instructors, thereby reducing the instructors' evaluation workload[4]. A remaining challenge of this approach is controlling the variability of subjective evaluations.

2.3 Hierarchical Structure among Workers

Extending the concept of peer review, Whiting et al. proposed an approach that organizes workers into a guild-like hierarchical structure[5], where peer evaluation is organized hierarchically. Their study demonstrated that such a hierarchical organization can mitigate variability observed in non-hierarchical evaluations, resulting in assessments that are closer to the ground truth. Moreover, studies of data work platforms in China suggest that guild-like hierarchical organizations help maintain workers' engagement and commitment[6]. The hierarchical structure within a guild may facilitate the effective capturing and sharing of requesters' subjective evaluation criteria within the organization. However, previous studies have only examined how closely the evaluations align with the ground truth in tasks that have unique true answers. Therefore, we should compare the hierarchical inter-worker evaluations with the requesters' subjective evaluations to examine whether the hierarchical approach is applicable to open-ended tasks as well.

3 Proposed Method

3.1 Hierarchical Inter-Worker Evaluation Structure

We construct a hierarchical inter-worker evaluation structure among workers, as shown in Fig. 1. This structure is constructed by classifying workers into levels according to their skills and is used for peer evaluation of their outputs on open-ended tasks. First, each worker starts at Level 1, and their level changes based on evaluations from requesters. Once each layer has enough workers, requesters evaluate only those at the highest level, thereby their workload remains low even as the system scales. Each worker evaluates those in the level below, enabling the requester's subjective evaluation criteria to propagate throughout the hierarchy. Furthermore, while not the focus of this study, we also expect that providing differentiated rewards based on levels will enhance worker engagement and commitment.

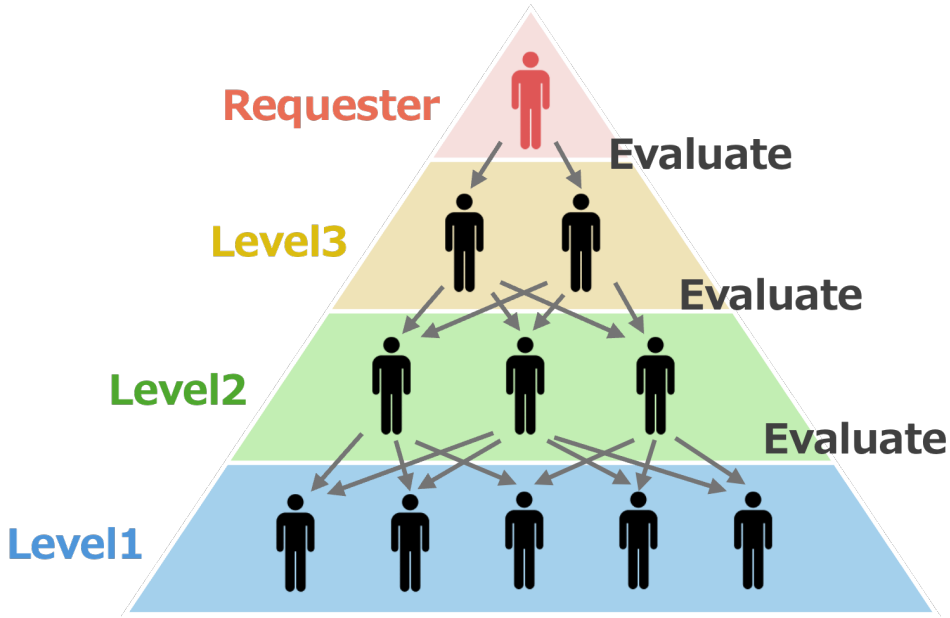


Fig. 1. Hierarchical Inter-Worker Evaluation Structure

3.2 Hypothesis

Based on previous studies suggesting that peer review can yield evaluations closely aligned with requesters' subjective evaluation criteria[4], and a hierarchical inter-worker evaluation is effective for evaluating workers on tasks with unique true answers[5], we propose the following hypothesis:

Evaluations performed within a hierarchical inter-worker evaluation structure can be closely aligned with the requesters' subjective evaluation in open-ended tasks.

4 Experiment

To test our hypothesis, we designed an experiment to examine how closely evaluations based on the hierarchical inter-worker evaluation structure align with those of the requesters.

4.1 Task Selection

As the open-ended task used in this experiment, we adopted an image captioning task based on the previous study by Aguirre et al[13]. Their study includes a process in which humans evaluate captions generated for visually impaired individuals. We adopted this task because the subjective human evaluation results are quantified using three criteria—fluency (how natural and readable the caption is), correctness (how closely the caption matches the facts in the image), and detail (how much information the caption contains)—which makes it easier to compare results numerically.

These three criteria were originally designed to evaluate captions intended for visually impaired individuals, for whom fluency, correctness, and detail are all simultaneously required. Because a deficiency in any of them can reduce the usefulness of a caption, we operationally define the unweighted average of these three criteria as an approximation of a worker's ability to produce captions aligned with the requesters' intent. The averaged

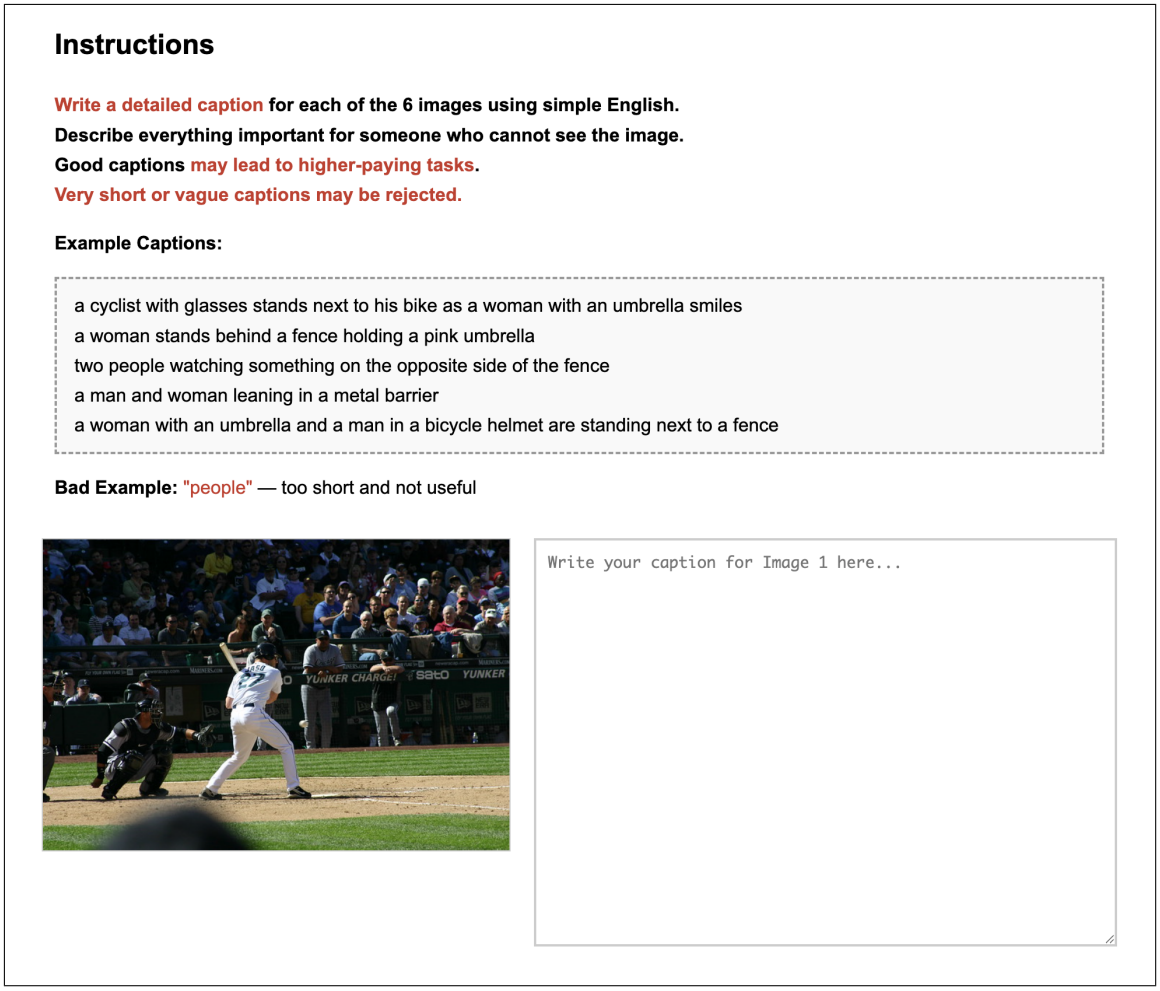


Fig. 2. Part of the Captioning Task (Captions are required for a total of six images.)

score (overall score) is used as an operational approximation of a worker's ability, rather than as a validated quality metric.

4.2 Procedure

First, we asked 30 workers from Amazon Mechanical Turk[3], a crowdsourcing platform, to generate captions for six images. Fig. 2 shows part of the captioning task. The six images were selected from the MS COCO captions dataset[4], following the previous study by Aguirre et al[13]. The reward for the captioning task was set at $0.50 per worker. Although the number of participants was limited due to evaluation workload constraints, this experimental setting was designed to serve as an initial validation of the hierarchical inter-worker evaluation structure.

Second, as requesters, we assigned three scores—fluency, correctness, and detail—to each of the six captions produced by every worker, on a scale of 0 to 10. Of the 30 workers, captions from 22 workers could be evaluated correctly; the remaining 8 workers were excluded due to duplicate answers or incorrectly completed fields. Then, the 22 workers were divided into three levels in descending order of the average of the three scores across

[3] https://www.mturk.com
[4] https://cocodataset.org

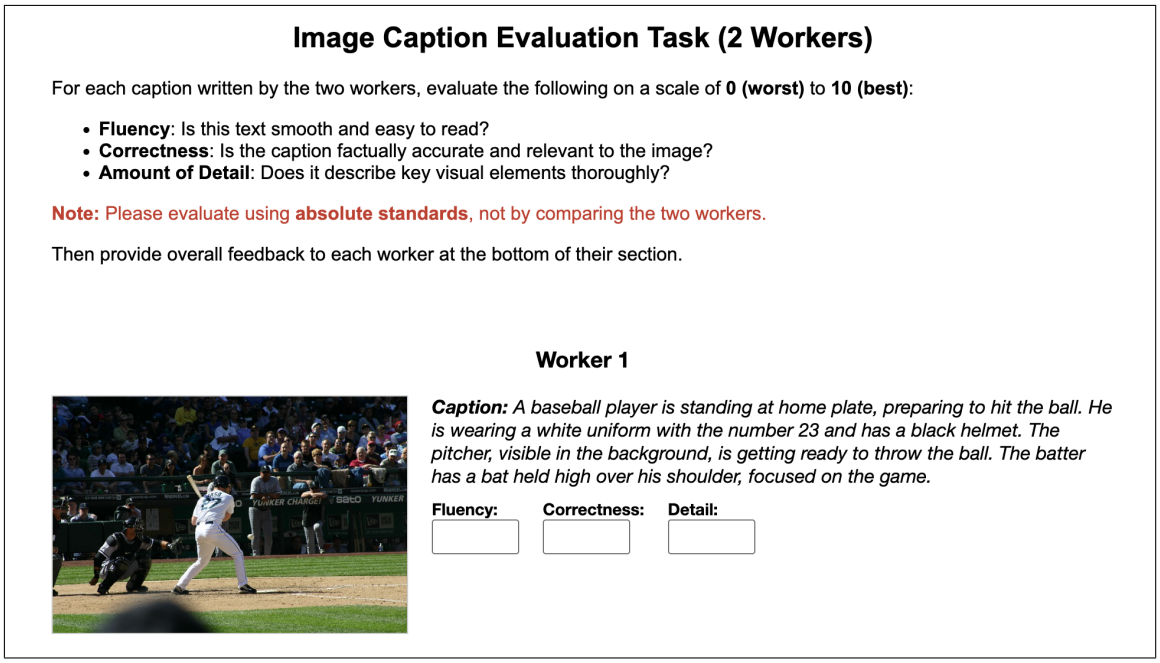


Fig. 3. Part of the Evaluation Task

all six image captions: the top 4 workers were assigned to L3, the next 6 to L2, and the remaining 12 to L1. This allocation ratio is intended to distribute evaluation tasks in the hierarchical structure.

Third, we prepared an evaluation task to evaluate image captions from L2 and L1 workers. Since L3 workers were supposed to be evaluated by requesters, their captions were not evaluated by other workers. That is, in this experiment, 18 workers at L1 or L2 were evaluated by other workers. The evaluation task was designed so that workers in L2 and L1 were each evaluated by two workers from the next higher level (L3→L2, L2→L1). The evaluation task included assigning three scores—fluency, correctness, and detail—to each of the six captions on a scale of 0 to 10, just as the requesters did. The reward for the evaluation task was set at $0.80 per worker. Fig. 3 shows part of the evaluation task. As a result, scores for 14 workers were collected. Since several workers did not perform evaluation tasks, scores for all 18 workers could not be collected as intended.

4.3 Analysis

For each worker, evaluations provided by two higher-level workers across the six images (12 evaluations in total) were averaged to obtain worker-level summary scores for fluency, correctness, and detail, as well as an overall score defined as the average of the three criteria. These averaged scores are intended to provide an approximation of each worker's ability to produce captions aligned with the requesters' intent, rather than evaluations of individual outputs, and are referred to as "scores from workers." Requester evaluations were averaged in the same manner to produce corresponding worker-level scores, referred to as "scores from requesters."

We calculated the following indicators between scores from workers and requesters: Mean Absolute Error (MAE), Mean Bias, Pearson's Correlation Coefficient (r), and Spearman Correlation Coefficient (ρ), to characterize their alignment from complementary perspectives. MAE quantifies the typical magnitude of differences in score values. Mean Bias captures whether scores from workers tend to be consistently higher or lower than scores

from requesters. Pearson's r indicates linear association between scores, reflecting proportional correspondence in score differences, while Spearman's ρ indicates consistency in the relative ranking of workers.

5 Results

This section presents the comparison results between the evaluation scores from workers and requesters. Table 1 shows indicators between scores from workers and requesters: MAE, Mean Bias, Pearson's (r), and Spearman's (ρ). MAE ranges from 1.446 to 1.869, with differences observed between criteria. The MAE for Correctness is relatively larger than that for the other three criteria. Mean Bias shows positive values across all criteria, indicating that scores from workers tend to be higher than those from requesters. Meanwhile, the magnitude of Mean Bias varied by criterion, with Detail exhibiting the smallest, and Correctness the largest. Pearson's (r) showed values of 0.8 or above for all criteria with Correctness showing a relatively lower value than the other three criteria. Spearman's (ρ) also showed values of 0.8 or above for all criteria. For Spearman's (ρ), the Overall criterion exceeded 0.9 and was the highest among the criteria, while Fluency showed the lowest value.

Table 2 shows the scores from workers and requesters for each worker in parallel. The workers are ordered in ascending order based on the Overall scores from requesters. Fig. 4 shows line graphs of scores for each evaluation criterion assigned by workers and requesters, plotted against the worker index. The worker index is ordered in ascending order based on the scores from requesters. Scores from workers exhibit a generally similar upward trend to those from requesters, with local variations. The figures visually show the tendency for scores from workers to be higher than those from requesters across criteria.

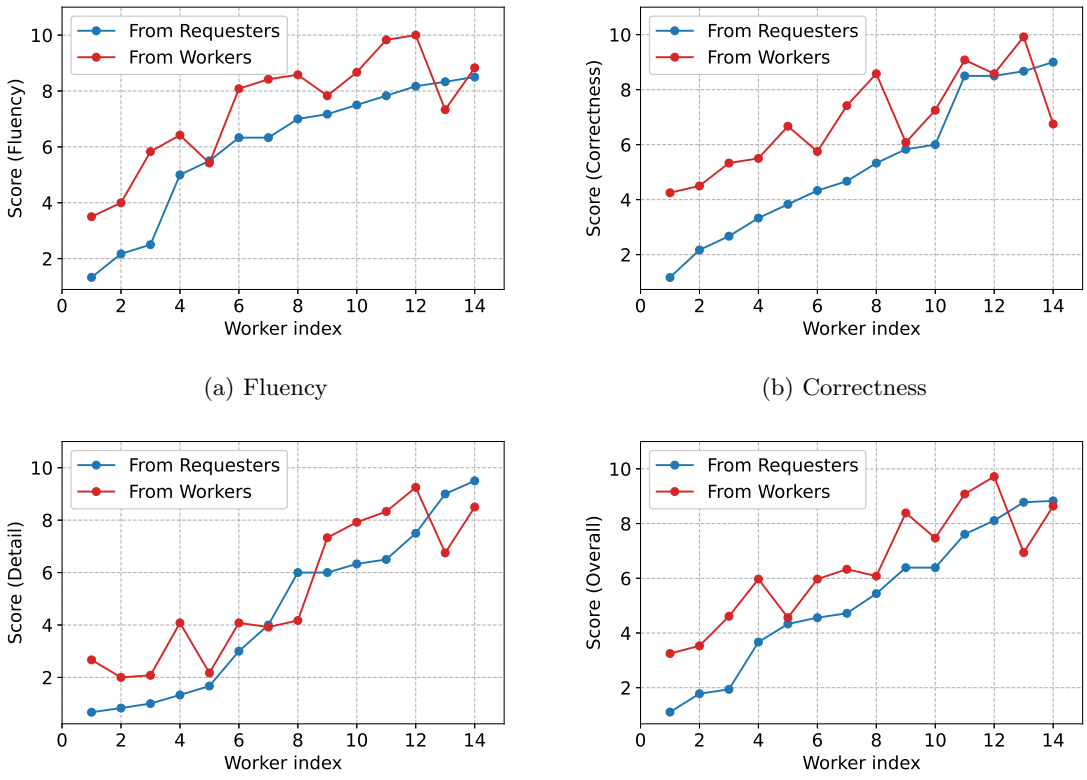
6 Discussion and Future Work

6.1 Discussion

The results indicate that, in terms of correlation, worker evaluations in the hierarchical structure exhibited a high degree of alignment with those of the requesters, reflecting strong consistency in the relative ranking of workers. However, in terms of absolute scores, evaluations from workers in the hierarchical structure tended to be higher than those from requesters, indicating differences in evaluation strictness.

The positive Mean Bias suggests that the reference points used for evaluation differed between workers and requesters, reflecting differences in evaluation perspective and purpose, as well as differences in how the rating scale was used. This finding contrasts with previous work suggesting that hierarchical evaluation structures can suppress workers' tendency to provide generous evaluations[5]. One possible explanation for this discrepancy is that, in our experimental setting, evaluations were not linked to promotion or rewards, which may have influenced how workers approached the evaluation tasks. As a result, differences in evaluation perspective or purpose between workers and requesters may have emerged.

Among the three criteria, Correctness and Fluency showed weaker alignment between evaluations from workers and requesters. For Correctness, larger differences in absolute scores indicate that workers and requesters may have applied different standards when judging correctness. For Fluency, although linear correlation was high, lower rank consistency suggests that workers and requesters differed in how they distinguished between captions of similar fluency. In contrast, the Detail criterion showed consistently higher



(a) Fluency

(b) Correctness

(c) Detail

(d) Overall

Fig. 4. Line Graphs of Scores from Requesters and Workers for Each Criterion

Table 1. Indicators between Scores from Workers and Requesters

Criterion	MAE	Mean Bias	Pearson r	Spearman ρ
Fluency	1.518	+1.363	0.892	0.814
Correctness	1.869	+1.548	0.826	0.834
Detail	**1.446**	**+0.708**	0.881	0.888
Overall	1.496	+1.206	**0.894**	**0.912**

Table 2. Scores from Workers and Requesters

(".w" and ".r" denote workers and requesters, respectively; "Correct" abbreviates "Correctness.")

Worker #	Fluency.w	Fluency.r	Correct.w	Correct.r	Detail.w	Detail.r	Overall.w	Overall.r
1	3.50	1.33	4.25	1.17	2.00	0.83	3.25	1.11
2	4.00	2.17	4.50	2.17	2.08	1.00	3.53	1.78
3	5.83	2.50	5.33	2.67	2.67	0.67	4.61	1.94
4	6.42	5.00	7.42	4.67	4.08	1.33	5.97	3.67
5	5.42	5.50	6.08	5.83	2.17	1.67	4.56	4.33
6	8.08	6.33	5.75	4.33	4.08	3.00	5.97	4.56
7	8.42	6.33	6.67	3.83	3.92	4.00	6.33	4.72
8	8.58	7.00	5.50	3.33	4.17	6.00	6.08	5.44
9	8.67	7.50	8.58	5.33	7.92	6.33	8.39	6.39
10	7.83	7.17	7.25	6.00	7.33	6.00	7.47	6.39
11	9.83	7.83	9.08	8.50	8.33	6.50	9.08	7.61
12	10.00	8.17	9.92	8.67	9.25	7.50	9.72	8.11
13	7.33	8.33	6.75	9.00	6.75	9.00	6.94	8.78
14	8.83	8.50	8.58	8.50	8.50	9.50	8.64	8.83

correlation and smaller Mean Bias. This suggests that criteria grounded in more explicitly observable aspects of the output, such as the amount of descriptive information, may facilitate more consistent evaluation even when subjective judgment by requesters is involved.

Interestingly, the Overall criterion exhibited the highest Pearson r and Spearman ρ. This suggests that aggregating multiple criteria may reduce criterion-specific noise and better approximate each individual worker's ability to produce outputs aligned with the requesters' subjective evaluation criteria. By aggregating multiple aspects of evaluation criteria, the Overall score may capture a more stable signal of relative performance than any single criterion alone. This finding has potential implications for the design of mechanisms that support worker development in line with the requesters' intentions. For example, if workers are grouped into relative performance tiers based on the Overall score, reward or feedback structures could be calibrated to these tiers, which may help guide workers' efforts toward further improvement.

As a limitation of this study, evaluation scores were obtained for only 14 workers, resulting in a smaller sample size than initially planned. Due to the limited number of workers evaluated from other workers (n=14), the results should be interpreted as exploratory. Notably, although several workers completed the captioning task, many did not participate in the evaluation task, highlighting the challenge of low worker retention in crowdsourcing. This observation suggests that, beyond designing effective worker evaluation and training mechanisms, we should also consider strategies that promote sustained worker engagement and retention.

6.2 Future Work

Future work will focus on strengthening the experimental validation of the hierarchical inter-worker evaluation structure. First, we plan to increase the number of workers to improve the statistical robustness and generalizability of the results. Second, because a substantial number of workers did not participate in the evaluation task, future work

should focus on improving worker engagement and retention. Sustaining long-term participation is essential for maintaining a stable pool of evaluators and for ensuring that trained workers continue to perform the tasks over time. Third, the tendency toward generous scoring observed in the experiment may stem from the experimental design, in which evaluation results were not linked to workers' promotion or rewards. Therefore, future work will explore evaluation mechanisms linked to promotion or incentives in order to encourage more responsible evaluations. Finally, by integrating these improvements: larger-scale experiments, engagement- and retention-aware task design, and incentive-linked evaluation mechanisms, we aim to conduct more comprehensive experiments to examine the effectiveness of inter-worker hierarchical evaluation structures for sharing requesters' tacit knowledge among workers.

7 Conclusion

In this study, we investigated how closely worker evaluations within a hierarchical structure align with requesters' subjective evaluations in open-ended tasks. The experimental results showed that workers' evaluations were highly aligned with the requesters' evaluations in terms of relative worker rankings. However, workers tended to evaluate more generously than requesters. Future work aims to design promotion and reward mechanisms to suppress generous scoring and enhance workers' skill development, engagement, and retention.

References

1. F. Daniel, P. Kucherbaev, C. Cappiello, B. Benatallah, and M. Allahbakhsh, "Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions," *ACM Comput. Surv.*, vol. 51, no. 1, Jan. 2018.
2. L. Chai, H. Sun, and Z. Wang, "An error consistency based approach to answer aggregation in open-ended crowdsourcing," *Inf. Sci.*, vol. 608, no. C, p. 1029–1044, Aug. 2022.
3. A. Chan, C. Di, J. Rupertus, G. D. Smith, V. Nagaraj Rao, M. Horta Ribeiro, and A. Monroy-Hernández, "Redefining research crowdsourcing: Incorporating human feedback with llm-powered digital twins: Incorporating human feedback with llm-powered digital twins," in *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '25. Association for Computing Machinery, 2025.
4. L. de Alfaro and M. Shavlovsky, "Crowdgrader: a tool for crowdsourcing the evaluation of homework assignments," in *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '14. Association for Computing Machinery, 2014, p. 415–420.
5. M. E. Whiting, D. Gamage, S. N. S. Gaikwad, A. Gilbee, S. Goyal, A. Ballav, D. Majeti, N. Chhibber, A. Richmond-Fuller, F. Vargus, T. S. Sarma, V. Chandrakanthan, T. Moura, M. H. Salih, G. Bayomi Tinoco Kalejaiye, A. Ginzberg, C. A. Mullings, Y. Dayan, K. Milland, H. Orefice, J. Regino, S. Parsi, K. Mainali, V. Sehgal, S. Matin, A. Sinha, R. Vaish, and M. S. Bernstein, "Crowd guilds: Worker-led reputation and feedback on crowdsourcing platforms," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17. Association for Computing Machinery, 2017, p. 1902–1913.
6. T. Yang and M. Miceli, ""guilds" as worker empowerment and control in a chinese data work platform," *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. CSCW2, Nov. 2024.
7. S. Dow, A. Kulkarni, S. Klemmer, and B. Hartmann, "Shepherding the crowd yields better work," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, ser. CSCW '12. Association for Computing Machinery, 2012, p. 1013–1022.
8. B. Saha Bhattacharya, B. Mandal, A. Biswas, and M. Bhattacharyya, "Improving character recognition by the crowd workers via corrective feedback," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 3982–3985.
9. W. Tang, M. Yin, and C.-J. Ho, "Leveraging peer communication to enhance crowdsourcing," in *The World Wide Web Conference*, ser. WWW '19. Association for Computing Machinery, 2019, p. 1794–1805.
10. C.-W. Chiang, A. Kasunic, and S. Savage, "Crowd coach: Peer coaching for crowd workers' skill growth," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, Nov. 2018.
11. D. Nicol, A. Thomson, and C. Breslin, "Rethinking feedback practices in higher education: a peer review perspective," *Assessment & Evaluation in Higher Education*, vol. 39, no. 1, pp. 102–122, 2014.
12. H. Zhu, S. P. Dow, R. E. Kraut, and A. Kittur, "Reviewing versus doing: learning and performance in crowd assessment," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '14. Association for Computing Machinery, 2014, p. 1445–1455.
13. C. A. Aguirre, A. Mahmood, and C.-M. Huang, "Crowdsourcing thumbnail captions via time-constrained methods," in *Proceedings of the 27th International Conference on Intelligent User Interfaces*, ser. IUI '22. Association for Computing Machinery, 2022, p. 36–48.

Authors

R Itano received a Master's degree in Engineering with a specialization in Computer Science from Doshisha University. He is currently pursuing his PhD in Engineering at Doshisha University. His research interests include human computation and crowdsourcing, and the development of training methods for short-term employment personnel.

H Tanitsu received her Bachelor's degree from the Faculty of Science and Engineering at Doshisha University. She is currently pursuing her Master's degree in Engineering at Doshisha University. Her research interests include human computation and crowdsourcing. She is working with Mr. Itano to establish effective training methods for short-term employment personnel.

M Bamba received his Bachelor's degree from the Faculty of Science and Engineering at Doshisha University. He is currently pursuing his Master's degree in Engineering at Doshisha University. His research interests include human computation and crowdsourcing. He is working with Mr. Itano to establish effective training methods for short-term employment personnel.

R Noseyama is a graduate student in the Faculty of Science and Engineering at Doshisha University. His research interests include human computation and crowdsourcing. He is working with Mr. Itano to establish effective training methods for short-term employment personnel.

A Kohiga is an associate professor in the Faculty of Science and Engineering at Doshisha University. His research interests include virtual reality, disaster-prevention IT systems, computer simulation, and AI coding.

T Koita is a professor in the Department of Information Systems Design, Faculty of Science and Engineering, at Doshisha University. His research interests include AI and machine learning, IoT, crowdsourcing, cloud computing, human computation, and UAV sensor networks.