

Crowdsourcing from the worker's perspective

Strategies to improve data quality and promote ethical platform use

Dylan Cole, Timothy Sullivan, Emma Zurek, and Simone Zhang

Abstract

Companies and researchers increasingly rely on crowdsourcing platforms for a variety of core tasks. This paper investigates how organizations can respect the dignity and needs of workers on these platforms, while also improving the quality of the data and insights that they gain. Focusing on the case of sustaining worker engagement in survey studies, we fielded a questionnaire that asked Amazon Mechanical Turk workers to share their perspectives on what influences their task completion and whether they return for follow-up studies. Results indicate that monetary incentives, reminder systems, task accessibility, and feelings of obligation and honor all enhance engagement. We conclude that researchers should ensure tasks are fairly compensated, craft tasks to be accessible and convenient, and take worker preferences and values into account when designing crowdsourcing tasks.

Introduction

Digital crowdsourcing platforms increasingly serve as the backbone of research and analytics activities across many sectors, with observers suggesting that they “look set to play a fundamental role in the future of commerce and society” (Grewal-Carr and Bates 2016). These platforms offer companies and other organizations ready access to large pools of research subjects and workers to quickly gather information about consumer preferences and habits, understand public attitudes, and annotate data for machine learning pipelines. One of the most prominent platforms is Amazon Mechanical Turk (MTurk). First launched in 2005, MTurk allows individuals and organizations (requesters) to post small-scale Human Intelligence Tasks (HITs) that its worldwide network of over 250,000 workers (MTurkers) (Robinson et al. 2019) can complete for compensation. The global reach of MTurk and similar platforms empowers researchers to cheaply and easily access more diverse and representative samples (Palmer and Strickland 2016), in theory strengthening data quality.

Yet despite the promise of platforms like MTurk, crowdsourcing platform workers, requesters, and observers have all raised a number of concerns about their use. A key ethical concern is that crowdsourcing platform workers regularly face poor, uncertain working conditions and receive inadequate pay. For instance, estimates suggest the median hourly wage for MTurk

workers was only \$2.00 per hour, with only 4% of workers earning more than the United States minimum wage (Hara et al. 2018). With 10% of MTurkers relying on crowdsourcing platforms to make ends meet (Williamson 2016), there has been increasing calls for new legislation to protect crowdsourcing workers (Cherry 2020).

Alongside these ethical issues, there are emerging concerns about the quality of the data companies and researchers gather from platforms like MTurk, with scholars drawing attention to issues like low-effort or false responses, low attention, and bots (Ahler et al. 2021, Moss 2021). Low data quality can undermine the validity and reliability of the information used to address important organizational and societal questions (Goodrich 2021).

How can companies and researchers that use crowdsourcing platforms ethically and productively engage workers on these platforms? In this white paper, we put this question to the workers themselves. We gather their perspectives through a survey that aims to understand how requesters can build a positive working relationship with workers, one that recognizes the contributions and dignity of workers and also improves data quality.

To get at this broader question, we specifically examine how to improve data quality and participant retention in multi-wave or longitudinal studies on MTurk, where requesters rely on the sustained engagement of MTurkers. We focus on this setting as it is where building a positive relationship between requesters and workers may be most important, and because longitudinal studies can yield particularly valuable insights, although their validity may be undermined when many participants drop out over time. Our survey suggests that organizations and researchers that turn to crowdsourcing platforms can build more productive relationships with workers by taking measures to ensure that tasks are accessible and convenient, and designed to fairly and fully recognize the value of workers' contributions.

Methods

We fielded a two-part pilot survey of U.S.-based MTurk workers. The survey aimed to capture the perspectives of people working on the platform: what are their work preferences and experiences? What suggestions do they offer for retaining participants over time?

Study population: Respondents were recruited via the CloudResearch service, which pre-vets study participants for high-quality responses. Only workers who had successfully completed more than 50 tasks on the platform and had an approval rate greater or equal to 95% were eligible to participate. We focus on this sub-population because it is a group commonly used in research and recommended by research on best practices (Peer et al. 2021).

Survey instrument: Part 1 of the survey collected information on respondents' demographics and background, as well as their MTurk work history, practices, and preferences.

Respondents were then asked to rate how likely they think they are to return for the follow-up on a seven-point scale, ranging from "1 - Very unlikely" to "7 - Very likely." Participants were then encouraged to participate in a follow-up survey to be posted the following day. Finally, they were asked: "What steps could we and other requesters take to better encourage people like you to participate in follow-up surveys?"

Part 2 was posted the following day. It featured questions about any surveys the respondent participated in in the preceding 24 hours, as well as an open-text question asking respondents why they decided to return for the follow-up.

Compensation: Respondents were paid \$1.50 for completing Part 1 and an additional \$1.00 for completing Part 2.

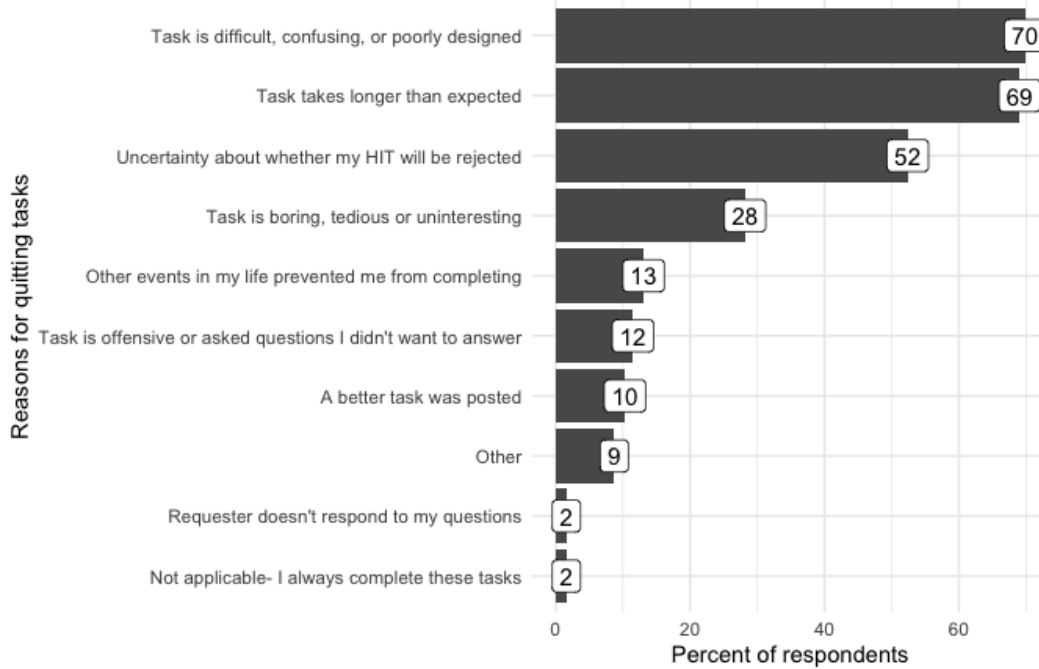
Analysis: We provide descriptive plots for key questions about survey completion and attrition. We also manually coded themes in open-text responses and computed how frequently different themes appeared.

Results

Part 1

Part 1 of our survey had 252 respondents. Overall, respondents indicated that they do not frequently quit tasks that they've started, with 38% stating that they never or rarely quit and 53% stating that they occasionally quit. When workers do quit tasks, however, it is most often due to tasks being "difficult, confusing, or poorly designed" or taking longer than expected, as illustrated in Figure 1. Over half of respondents further indicated that they were more likely to quit if they were unsure whether their work would be rejected and go unpaid.

Figure 1. Top 3 factors that most contribute to task quitting



When asked to rate how likely they believe they would be to participate in Part 2 of our survey, nearly all respondents – 96% – indicated that they were “likely” or “very likely” to participate in the follow-up survey, suggesting that most *intended to* continue their participation.

We then also asked respondents to share what steps we and other requesters could take to better encourage people like themselves to participate in follow-up surveys. Table I summarizes the prevalence of common suggestions.

Table I. Top MTurk worker suggestions for sustaining participation

Suggestion	Percent of Respondents
Send reminders (email, text, etc.)	43%
Increase overall monetary incentives	35%
Provide bonus/higher pay for follow-up participation	30%
Offer flexible response windows	17%
Increase task transparency	9%

Better design tasks	6%
Other	15%

Note: N = 252

The most common suggestion, cited by 43% of participants, was to set up a reminder system when new tasks or follow-up surveys are posted. One participant stated, for instance, that “I would say there's a near 100% chance of me participating in a follow up if I got a[n] email about the follow up being posted.”

Monetary incentives and a bonus/higher pay for completing later components were also popular suggestions with 35% and 30% of respondents suggesting them. For instance, One respondent endorsed increased pay, saying, “The more you pay, the more likely people will show up to take the follow up.”

Flexible response windows were important for 17% of respondents. In the words of one respondent, “Ensure that the survey is available for a long period of time since participants might be busy with other life responsibilities.”

Another common suggestion was to improve task transparency. Detail was key for these respondents, with one sharing, “The requester needs to be transparent about how long the study will take and how much I will be paid ... They need to declare how the data will be used and not ask questions of a personally identifying nature.” Respondents also indicated additional factors, such as whether an accurate estimate of the time necessary to complete a survey is provided, whether the benefits of responding are clear, and whether a survey is well-designed.

Part II

Of our initial sample, 215 respondents (86%) returned to complete Part II. Table 2 summarizes the most common reasons given.

Table II. Top reasons for returning

Reason	Percent of Respondents
Monetary Incentive	62%
Honor/Obligation/Pride	31%

Simplicity	15%
Convenience	13%
Received reminder	11%
For the good of the research	11%

Note: $N = 215$. In free-text responses, respondents often included multiple rationales per submission.

Perhaps unsurprisingly, the most common reason respondents gave for their continued participation was the monetary incentive the follow-up survey provided. Many participants indicated that they returned to “earn money” and that “the payment is great”, with 62% citing this reason.

However, the second most common reason, cited by 31% of respondents, was not about external rewards. Instead, it was related to respondents feeling a sense of honor or obligation to the researcher or wanting to take pride in their work. One respondent captured these feelings by sharing that “I feel obligated to finish what I started. I want to ensure that the survey givers have an adequate amount of responses to help with their research.” Indeed, many participants, having indicated that they would likely return for the follow-up, stated that they returned to “try and keep [their] word.” Like obligation, many respondents also indicated that they responded for the “good of research,” with 11% of respondents sharing that rationale.

Finally, survey simplicity, accessibility, and reminder systems were also important factors for around 1 in 10 respondents.

Implications for Practice

Our results reveal some promising avenues by which requesters could increase data quality, improve follow-up response rates, and enhance the experiences of crowdsourcing workers.

Pay well: Respondents consistently cited pay as a significant factor that influences whether they stay engaged, reflecting the reality that the platform is an important source of income for nearly one-third of American MTurkers (Williamson 2016). Thus, our results suggest that increasing pay – or likewise offering compensation bonuses – will increase follow-up participation. Requesters should carefully test surveys and other tasks to get an accurate sense of how long they take and set fair wages accordingly. Moreover, past research suggests that better pay encourages participants to share how they actually feel, rather than give the easy middle answer (Goldstein 2018).

Meet people where they are and lower participation costs: Our results also suggest a need to make survey participation more accessible through reminder systems, flexible response windows, and better designed surveys. Many workers intend to continue participating, but life can get in the way. Some 13% of respondents in this study said that other events in their lives kept them from fully completing tasks. By taking steps to accommodate the conflicting commitments of MTurkers, requesters may be able to increase worker volition – feelings of self-direction over their employment (Philip and Davis 2020)– to attain not only higher quality data but also build more productive relationships with workers. Additionally, since 70% of MTurkers who have quit tasks indicated they tend to do so when tasks are difficult, confusing, poorly designed, or take longer than expected, requesters should make sure to conduct small-scale pilots and usability tests to reduce technical frustrations and confusing directions.

Recognize the dignity of workers: Our results offer a reminder that for many MTurkers, their work is a source of pride and dignity. Although they operate largely on an anonymous platform, MTurk workers can feel a positive sense of connection to requesters. They are more likely to remain engaged when they are treated with respect and feel that their contributions are fairly compensated and serve a greater purpose. Our findings further suggest that it may be valuable to ask participants about their intent to return for future surveys and to solicit their feedback about how to improve follow-up participation. The former may act as an implicit commitment device while the latter can help give workers voice and surface potential study-specific actions that requesters can take to improve retention.

Increase task transparency: Greater transparency can demonstrate to MTurk workers that their contributions are valued. Our findings align with past research, which has shown that transparency concerning the aims and benefits of a study, especially how it may serve a wider community, can boost participation and response quality (Saleh and Bista 2017:70). Additionally, clear expectations about data quality are important, as more than half of respondents indicated that they quit studies when they are uncertain whether their submissions will be accepted. Respect for privacy is also important to some MTurkers, with 12% of respondents indicating that they have quit tasks because of a desire to not disclose certain information. Accordingly, requesters should clearly state upfront what they will ask of participants and describe how participants' data will be used in accessible language.

Conclusion

Amazon Mechanical Turk is a popular platform for companies and researchers to obtain data more cheaply and efficiently than traditional methods. The platform continues to grow: during the initial wave of COVID-19 lockdown measures, a surge of new workers joined the platform's user base. This expansion has the potential to increase the diversity of the worker pool, but scholars have observed an uptick in data quality challenges, as “post-quarantine

workers are also more likely to be inconsistent, fail attention checks, and give responses ... that are neither intuitive nor correct” (Arechar and Rand 2021:2594). Therefore, it is more important than ever to determine what requesters can do to consider the needs of crowdsourcing platform workers to improve worker experiences and assure high data quality.

Our study highlighted these needs by capturing the voices of workers themselves, asking them to identify the factors that, in their view, most dampen or bolster continued engagement. We find that traditional strategies such as greater monetary incentives and sending reminders shape MTurkers’ decision-making. At the same time, considerable proportions of respondents also reported their continued engagement is shaped by self-conceptions of obligation and honor, as well as their personal interest in the survey topic and the sense that they’re contributing to important research. Thus, these findings highlight the importance of ensuring that work on crowdsourcing platforms is designed to be fairly compensated, accessible, engaging, transparent about how data will be used, and clear in its social value.

References

Ahler, Douglas J., Carolyn E. Roush, and Guarav Sood. (2021). The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. *Political Science Research and Methods*, 1–20. <https://doi.org/10.1017/psrm.2021.57>

Arechar, Antonio A. and David G. Rand. (2021). Turking in the time of COVID. *Behavior Research Methods*, 53(6), 2591–2595. <https://doi.org/10.3758/s13428-021-01588-4>

Cherry, Miriam A. (2020). A Global System of Work, A Global System of Regulation?: Crowdwork and Conflicts of Law (SSRN Scholarly Paper No. 3523303). <https://papers.ssrn.com/abstract=3523303>

Goldstein, Markus. (2018, February 21). If you pay your survey respondents, you just might get a different answer. *World Bank Blogs*. <https://blogs.worldbank.org/impactevaluations/if-you-pay-your-survey-respondents-you-just-might-get-different-answer>

Goodrich, Erin R. (2021). Six Reasons Why Research is Important. *Association of Internet Research Specialists*. <https://aofirs.org/articles/six-reasons-why-research-is-important>

Grewal-Carr, Vimi and Carl Bates. (2016). The three billion: Enterprise crowdsourcing and the growing fragmentation of work. *Deloitte*. [https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/us-cons-enterprise-crowdsourcing-and-growing-fragmentation-of-work%20\(3\).pdf](https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/us-cons-enterprise-crowdsourcing-and-growing-fragmentation-of-work%20(3).pdf)

Hara, Kotaro, 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. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3173574.3174023>

Moss, Aaron. (2021). The Truth About Online Data Quality: 30% of Survey Respondents Live on Pluto. *CloudResearch*. <https://www.cloudresearch.com/resources/blog/the-truth-about-online-data-quality/>

Palmer, Joshua C. and Justin Strickland. (2016). A beginner's guide to crowdsourcing. *Psychological Science Agenda*. <https://www.apa.org/science/about/psa/2016/06/changing-minds>

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01694-3>

Philip, Jestine, and Mark A. Davis. (2020). Presenting Amazon's Mechanical Turk as more than just a data sample – A study of MTurker experiences. *The Irish Journal of Management*, 39(2), 85–98. <https://doi.org/10.2478/ijm-2020-0006>

Robinson, Jonathan, Cheskie Rosenzweig, Aaron J. Moss, and Leib Litman. (2019). Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PLOS ONE*, 14(12), e0226394. <https://doi.org/10.1371/journal.pone.0226394>

Saleh, Amany and Krishna Bista. (2017). Examining Factors Impacting Online Survey Response Rates in Educational Research: Perceptions of Graduate Students. *Journal of MultiDisciplinary Evaluation*, 13(29), 63–74.

Williamson, Vanessa. (2016). Can crowdsourcing be ethical? The Brookings Institution. <https://www.brookings.edu/blog/brown-center-chalkboard/2016/02/08/can-crowdsourcing-be-ethical>