# Student Notebook: Benefits, Drawbacks, and Suggestions for Using Amazon's Mechanical Turk

January 30, 2019

Amazon's Mechanical Turk (MTurk) is an online crowdsourcing platform designed to aid in recruiting people to complete various tasks (Buhrmester, Kwang, & Gosling, 2011). Overall, Amazon advertises its MTurk service as offering access to over 500,000 different workers from 190 countries; however, the majority (more than 75%) of MTurk workers live in the United States and India (Paolacci & Chandler, 2014). The tasks posted on MTurk by "requestors," referred to as human intelligence tasks (HITs), range in length and duration and are completed by "workers" for a set, usually small, fee. Tasks posted by requesters on MTurk are referred to as human intelligence tasks (HITs).

MTurk is a great data collection tool for graduate student researchers who are investigating a novel trend but might be concerned with finding large amounts of participants in a reasonable amount of time. MTurk can also be helpful for someone trying to expand the generalizability of their project from the typical research conducted using a predominantly Caucasian/European-American, affluent, undergraduate population. While MTurk can be beneficial for gathering a diverse sample in an abbreviated length of time, certain drawbacks should be considered when using this crowdsourcing service. Below are several benefits and drawbacks to using MTurk for data collection.

**Benefits:**

- Overall, the sample collected from MTurk is likely to be more diverse than a sample of undergraduate students (Buhrmester et al., 2011). Participants are generally older, more geographically representative of the US, and more diverse than participants collected from undergraduate samples.
- The reliability of data collected from MTurk has not been found to be significantly different than data collected by other means. Participants who respond using MTurk generally answer reliably and consistently, as evidenced by high test-retest reliability rates even after a period of 3 weeks (Buhrmester at al., 2011).
- MTurk software supports the embedding of other survey software (e.g., Qualtrics). In this regard, many different types of research methodology are possible using MTurk workers, including longitudinal, qualitative, and mixed methods.

**Drawbacks:**

- Research shows that users of MTurk have some fundamental differences from the general population. MTurk workers are more educated, less religious, and more likely to be unemployed than the general population (Goodman, Cryder, & Cheema, 2013). If a researcher is trying to investigate specific trends within minority populations, such as levels of religiosity, or educational differences, these cultural differences could confound future results and limit generalizability.

- The range of ages and socioeconomic statuses of MTurk workers could be more limited than those found in the general population. While MTurk appears to include a diverse sample of workers, logically, older adults might be less likely to utilize technology. Fundamentally, MTurk requires the usage of some web-based platform along with the availability of the technology to accommodate such activities (e.g., a computer, a laptop, an iPad). With older adults and those within lower socioeconomic statuses, many might not have access to the technology needed to use MTurk. Additionally, particularly with older adults, there might be a lack of familiarity with web-based services such as MTurk, leading to a lower likelihood of use.
- Diversity is not synonymous with representativeness. Research suggests that the amount of workers using MTurk who belong to certain racial/ethnic groups might be lower than the amount found in the general population (Paolacci et al., 2014). Particularly, this trend has been found relative to African American and Hispanic American workers (Paolacci et al., 2014).

Given the above limitations, when sampling workers in MTurk you may be most likely to encounter Caucasian, technologically-adept, highly educated secular workers. Several helpful strategies exist, however, to mitigate these drawbacks and obtain your desired sample.

**Suggestions:**

- *Be very explicit in your HIT title and description.* Though MTurk has the capability for researchers to purchase "qualifications" that parcel out groups of people according to certain specifications, as of yet there is no "qualification" specifically for demographics, such as race and ethnicity. To control for this limitation, in both the title and description of the HIT, use uppercase letters for the demographic specifications of interest. This method can streamline the process and help gather many more participants from the population of interest.
- *Implement "checks" into your task that assess the demographics of the person responding.* An additional method to collect responses from participants consistent with the specification of interest is to include a "check" into your task. The participant should fill this "check" out before they begin the actual task. For example, in a study that I was working on, participants who were not African American were still submitting responses even though the title and description for the HIT explicitly indicated the desire for solely African American participants. To reduce the potential for these responses, we added a question before the administration of the research questionnaire asking, "What is your race/ethnicity?" In this way, we separated out those who had gotten through to the study who did not meet the demographic qualifications of interest.
- *Understand and accept that recruiting diverse populations through MTurk might be a slow process.* One of the advantages of MTurk is the ability to recruit a large number of participants in a relatively inexpensive, expedited manner (Follmer, Sperling, & Suen, 2017). It is important to remember, however, that the majority of MTurk users are Caucasian/European American. Therefore, if you are attempting to sample for participants from a specific minority group, you need to be persistent to collect a large sample. Often, researchers who are able to collect their data more quickly may not be seeking to gather participants from a specific minority group.

MTurk can be a great means of recruiting a diverse sample quickly and in a cost-efficient manner; however, the inherent differences observed between an MTurk sample and a sample collected using traditional methods might present significant challenges in generalizing the results of the study. These differences include faith-based, technological, educational, age-related, socioeconomic, and employment-

related differences. Additionally, the same ethical guidelines that you would uphold with participants collected from any other population must be maintained with MTurk workers despite the limits this program places on personal interaction. Always be mindful of the implications of using MTurk, and good luck with data collection.

**References**

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, 6(1), 3–5. doi:10.1177/1745691610393980

Follmer, D. J., Sperling, R. A., & Suen, H. K. (2017). The role of MTurk in education research: Advantages, issues, and future directions. *Educational Researcher*, 46, 329–334. doi:10.3102/0013189X17725519

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26, 213–224. doi:10.1002/bdm.1753

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184–188.