Utilizing Amazon Mechanical Turk to Gather Data: Pros and Cons
Michael Obal, Doctoral Student, Temple University

Amazon Mechanical Turk, commonly referred to as MTurk, has recently become a popular tool for gathering data amongst researchers and practitioners alike. MTurk is a crowdsourcing Internet marketplace that allows users to post and complete tasks, known as Human Intelligence Tasks or HITs, on their site. Requesters post a HIT on the site that they would like completed alongside a payment for completing that task and the number of HITs they would like completed (in other words, the desired sample size). On the other end, workers can browse through posted HITs and choose to complete HITs that appear interesting to them and offer a decent enough pay. After a worker has completed a task, the requester can choose whether or not to accept the task and therefore whether or not to pay. If a worker does not satisfactorily complete the task, their approval rating will decrease. In turn, the workers can provide feedback about the requester that will help other workers evaluate the requester. This worker-requester evaluation system serves as the primary feedback mechanism on MTurk as Amazon does not monitor the service.

For social scientists and marketing practitioners, MTurk is extremely attractive because of its pricing structure, ease-of-use, and time to completion. Requesters can set and adjust their own prices as they see fit. In general, the cost of a worker on MTurk is substantially cheaper than using a professional paneling service. Services such as Qualtrics or Survey Monkey commonly charge a researcher $5-$10 for the completion of a ten-minute survey. In contrast, a higher-end worker on MTurk can expect to make $6-$12 per hour, which means the requester should only have to pay $1-$2 for the completion of a ten-minute survey. Despite the lower pay rate, finding willing survey takers is not a major issue. A requester that is offering $1.00 for the completion of 50 ten-minute surveys should expect to have that request filled within a few days. Conveniently enough, this can all be done at home and processed using the requester's Amazon account.

However, while the ease of attracting workers is appealing, an obvious question arises from MTurk: are these worker responses legitimate? Surprisingly, the answer is: generally, yes. Researchers at Princeton, NYU and Ca'Foscari University of Venice conducted an experimental study study in 2010 that compared responses from MTurk users to responses from students at a major Midwestern university. Interestingly, they found no significant differences between the responses offered by the MTurk users compared to the students. They also found that the non-response error rates were quite similar, as approximately 92% of MTurk users completed the study compared to about 98% of students. Furthermore, MTurk includes a few filters that can help requesters improve non-response rates, such as a requirement that workers must meet an approval rating threshold in order to view your HIT.

Further findings have suggested that there may even be some advantages to using an MTurk sample over a student sample. Because MTurk workers can complete the tasks anonymously and without contact with the researchers, internal validity can be improved as methodological issues such as experimenter bias and subject crosstalk can be minimized or even completely eliminated. Furthermore, a sample drawn from MTurk is generally more representative of the general population than a student sample, a common limitation for any study utilizing an undergraduate population. A 2012 study from researchers at MIT, Yale and Cal-Berkeley found that the MTurk population is actually a relatively close match to the U.S. population. Their sample of MTurk users was comparable to national averages on gender splits, education levels, race and median age – all measures in which undergraduate samples fall short. Because of this, MTurk should be considered a desirable outlet for any researcher looking for a random sample of the general population.

However, the limitations of MTurk become more apparent if a more precisely defined population is desired. MTurk does not offer requesters the opportunity to define their populations in the same manner that many professional paneling services do. Furthermore, any attempts to define the population through MTurk may be skirted by advanced users. As discussed at this year's Association for Consumer Research Conference in Vancouver, MTurk users are not naïve. In fact, they are actually quite savvy. Any attempt to restrict MTurk users (e.g. "This survey is only for respondents under 30 years old.") will be met with false responses from MTurk users. MTurk users are driven to find appealing tasks and, since their responses are anonymous, there generally is no punishment for falsifying information in order to qualify to complete a task.

Because of this issue, requesters would do best to not filter respondents at all and instead simply use the responses they need for analysis after data has been collected. For example, a forthcoming article of mine in the Journal of Service Management utilized an MTurk sample to look at how Baby Boomers viewed a website as compared to Millennials (e.g. Gen Yer's). As such, I had little use for respondents who fit into the Gen X category. However, I could not have asked Gen Xer's to just ignore my study; they would have simply lied about their age and completed the study, thus invalidating my results. Instead, I collected results for everyone and asked them to indicate the age group they belonged to (Note: researchers should also avoid asking for specific ages, as most people have a tendency to round down! Asking for a respondent's age group will yield more accurate results). I then grouped the Millennials and Baby Boomers together and compared their responses and ignored everyone else. I still have plenty of data on Gen Xer's, but this age group did not fit into the theoretical framework of my paper and therefore their responses will probably just remain in a data file on my computer without ever seeing the day of light. Fortunately, as noted before, costs are less of an issue with MTurk, so I didn't feel too bad about paying for extra data.

In sum, while an MTurk sample may have a few issues, a clever researcher may find it to be quite useful. In order to ensure that your sample is of a high quality, I recommend the following. First, only use MTurk if you are seeking a general population or something close to it. The more defined your population is, the less desirable an MTurk sample will be. Second, don't assume that the MTurk users are naïve. MTurk users are often very experienced study subjects and survey takers, so it is best to avoid classic manipulations and minimize the clues that may hint at the purpose of your study. Third, save the workers' IDs so that you may delete any repeat workers. Fourth, encourage MTurk users to take their time and eliminate those who complete the task too quickly from your final sample. Fifth, add a few extremely easy questions (e.g. "An octopus has how many arms?") and eliminate respondents who answer incorrectly. Sixth, run your study through an external survey provider and require MTurk users to enter a study completion code that can only be accessed once the external survey has been completed. By doing this, the researcher is able to eliminate those users who do not pay attention or attempt to get paid without completing the survey. Seventh, scan your final results for any suspicious responses (e.g. respondent gives the same answer for all questions). MTurk users are essentially hourly employees, which means they will have an innate desire to complete studies quickly. Be on the lookout for these types of users. Finally, make your task as interesting to complete as possible! Previous interviews with MTurk users have indicated that many of them complete tasks for fun. Therefore, if you can create an exciting study for them to complete, then you will be more likely to gather legitimate data.