

## Online data collection systems: MTurk and beyond

Barbara Nevicka

Bruno Verschuere

Online data collection systems have broadened the scope of data collection in behavioral research. Data sets that used to take months to compile can now be collected in a single afternoon. This has led to an increase in the frequency and proportion of such studies being used in behavioral research. Amazon's Mechanical Turk (MTurk) platform was one of the first data collection system in which registered users from around the world, or workers, complete surveys and computerized tasks, referred to as Human Intelligence Tasks or HITs, for small financial incentives. By now there are several alternative platforms, such as Figure Eight (formerly Crowdfunder) and Prolific.

Below we highlight some potential advantages and disadvantages of such data collection strategies.

### Potential advantages

- Quick, easy and inexpensive access to online participants
- This efficiency in data collection enables attainment of large samples sizes to increase statistical power and to easily conduct follow-up and/or replication studies
- MTurk reaches a more diverse population than the typical student sample and even some community samples, allowing researchers to gain generalizability to broader populations and to test research questions across different cultures
- Allows selecting subsamples depending on research needs (e.g., only smokers; only native English speakers; or only those who have full-time employment).
- Research has generally suggested that MTurk workers provide high-quality data, demonstrating psychometric equivalence to other data collection methods with respect to internal consistency and test–retest reliability (not just for surveys but also for RT-research)
- MTurk workers seem to be attentive to study instructions and tasks. They demonstrate passing rates on instructional manipulation checks (IMC) that are typically comparable to other data sources (e.g., community samples, other online samples). One notable exception is that MTurk samples were found to perform worse on IMCs as compared with an undergraduate subject pool.
- In a recent study, MTurk workers did not differ meaningfully from the community normative sample on any of the four validity scales, offering further support for the general consensus that MTurk workers provide high-quality data and are generally attentive to study tasks.

- The presence of overreporting or underreporting of clinical symptomatology does not appear to vary from what might be expected from a large, representative sample.
- Many commonalities have been found between MTurk participants and traditional samples, contributing to the growing literature showing that MTurk participants give responses similar to other traditionally used samples (e.g., showing similar judgment and decision biases, such as framing effects)

### **Potential disadvantages**

- With some exceptions (e.g., Prolific), platforms were typically not developed to conduct scientific research and do not provide support for that purpose. There are now platforms that assist with online research (e.g., Turkprime facilitates using MTurk for research purposes).
- Non-naiveté, e.g., the typical MTurker has engaged 5 times in the trolley paradigm (a popular research paradigm for moral decision making). You can, however, choose to exclude high active workers.
- Ethical dilemmas with regard to payment. You may pay peanuts and still get participants. That is not necessarily a problem (e.g., when participation out of interest or to help science), but can evoke ethical dilemma when participants come from low-income countries and participate as a (part-time) job.
- The platform population may be much smaller than what is advertised. E.g., while there are millions of MTurkers, a typical laboratory can access about 7,300 workers. On the positive side: The estimated time taken for half of the workers to leave the MTurk pool and be replaced is about 7 months.
- For RT research you can use existing software (e.g., Web version of Inquisit), but that typically requires downloading plugins which can deter workers. Tasks in Javascript are more user-friendly but require more programming skills.
- The relatively low financial compensation that they receive provides an incentive for workers to complete tasks as quickly as possible in order to maximize earnings, potentially at a cost to accurate responding. Although, their attitude to money and time is more similar to student participants rather than community participants.
- MTurk samples are not representative of the broader population:
  - MTurk participants are less extraverted, less agreeable, more neurotic and have lower self-esteem than other participants, which could present challenges for some research domains.
  - MTurk workers seem to have difficulty with social connectedness, reflected by reports of social isolation, limited social support, social detachment and interpersonal coldness and resentment.
  - MTurk workers seem to score higher on social anxiety, autism spectrum disorder and negative affect.

- MTurk participants have attitudes about money that are different from a community sample's attitudes but similar to students' attitudes
- MTurk participants are more likely to use the Internet to find answers to factual questions, even with no incentive for correct responses.

### **Tips for using MTurk (or similar) samples**

Depending on the research question, online data collection systems can offer valuable opportunities for data collection. Below are some tips to consider to ensure data quality:

1. Include screening questions that gauge attention and language comprehension. Filtering participants by whether they correctly answered the IMC or not was found to reduce statistical noise.
2. Making explicit that the study is only accessible for certain users (e.g., 'Study only for Smokers!') increases cheating.
3. Note that MTurk may not be appropriate for long or complicated studies in which participants may be more likely to lose attention and not follow instructions. Previous research found that MTurk participants were equally attentive as other participants when the study was short (~5 minutes).
4. One way to mitigate the lower attention associated with MTurk samples could be to emphasize the scientific importance of the study to participants, encourage them to be attentive, or work with extra rewards (e.g., pay for successful completion of the task).
5. Use screening features (e.g., with TurkPrime, which is a technology company that optimizes participant recruitment platforms (such as MTurk) for scientific research) to select only high quality "workers" as determined by their prior participation attempts.
6. Avoid questions with factual answers, or if you cannot avoid such questions provide MTurk workers with specific instructions to not use external resources. This has been found to significantly reduce such occurrences of "cheating" on factual questions.
7. Consider how individual differences in financial and social domains may influence your results.
8. Consider providing compensation comparable to that afforded to traditional participants, as underpayment may impact data quality.

## References

- Anderson, C. A., Allen, J. J., Plante, C., Quigley-McBride, A., Lovett, A., & Rokkum, J. N. (2018). The MTurkification of social and personality psychology. *Personality and Social Psychology bulletin*.
- 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.
- McCredie, M. N., & Morey, L. C. (2018). Who Are the Turkers? A Characterization of MTurk Workers Using the Personality Assessment Inventory. *Assessment*.