

Taking the Plunge: Online Data Collection Using MTurk

L. Jean Harrison-Walker, walker@uhcl.edu

Abstract

Mechanical Turk, an online crowdsourcing service created in 2005 by Amazon.com, is gaining in popularity among academics as a source of subjects for survey research. The initial attraction is based on access to a very large pool of participants who can be accessed online producing rapid response at low cost with minimal effort on the part of the researcher. For the uninitiated, however, the task of designing and running a research project on MTurk is rather daunting.

This paper begins with an introduction to MTurk. We review the language of MTurk and explain how a researcher goes about setting up an online survey. Next we talk about several of the decisions that have to be made by the researcher. Decisions relate to the compensation rate to be paid, qualifications for respondents, timing decisions, and aspects of project design to aid in filtering out bad data. Such decisions are especially complicated when it comes to lengthy surveys.

In the third section of this paper, we review the concerns associated with survey data collection using MTurk. These include concerns about (1) generalizability, (2) replication of findings, (3) the effect of reward on data quality, (4) the growing number of workers from India, (5) study length and (6) inattentiveness.

The fourth section evaluates some of the key implementation decisions that must be made by researchers. These decisions involve (1) the inclusion, format and placement of attention check questions; (2) the amount of the Reward, (3) the minimum level of Respondent experience and approval rating, (4) the estimated time to complete the survey as well as the time by which to submit the HIT, (5) posting the HIT, and (6) the criteria to be used to reject submissions.

Next, we present batch-level metrics for our data collection using a lengthy survey. Data was collected using a 143 item survey about service failure recovery designed on Survey Monkey and linked to MTurk. The target sample size was set at 600 Respondents. The estimated time required to complete the survey was 45 minutes; however Respondents were allowed a total of 90 minutes to return from Survey Monkey to MTurk and enter their identification code. Eight reasons for rejection were used during the survey approval process. In addition to the one pretest (N=5), data was collected in 13 separate batches. The decisions made for each batch are

identified and explained. The batch level metrics that are reported and discussed include data includes the date of posting, respondent approval rating, the number of hits requested, the number of hits received, the number of surveys accepted, and the percent of surveys accepted. We present data regarding the number of surveys rejected based on each of the eight rejection criteria.

Across the data collection, 938 HITS were submitted and 605 were approved, producing an average acceptance rate of 64.5%. In addition to the 938 HITS that could be processed for either acceptance or rejection of payment, there were an additional 66 Respondents who could not be processed. These were people who went to the Survey Money link but failed to return to MTurk and therefore did not show up in the MTurk results to be processed for payment. Some Respondents answered a single question before opting out of the survey, while others completed one of the two electronic pages before exiting the survey. If a Respondent does not go back to MTurk, their Worker number is never recorded by MTurk.

We conclude by identifying a number of findings from our MTurk data collection experience that were particularly notable, if not downright surprising. Recommendations are made for various decision areas to promote a more successful MTurk experience for future researchers.

References

Buhrmester, M., Kwang, T. & Gosling, S.D. (2011) Amazon's Mechanical Turk: A New Source of Cheap, Yet High Quality, Data? *Perspectives on Psychological Science* 6, p. 3-5.

Fort, Karen, Adda, Gilles & Cohen, K. Bretonnel (2011) Amazon Mechanical Turk: Gold Mine or Coal Mine? *Computational Linguistics* 37(2), p. 413-420.

Goodman, Joseph K., Cryder, Cynthia E. & Cheema, Amar (2013) Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples, *Journal of Behavioral Decision Making* 26, p. 213-224.

Horton, J.J., Rand, D.J., & Zeckhauser, R.J. (2011) The Online Laboratory: Conducting Experiments in a Real Labor Market, *Experimental Economics* 14, p. 399-425.

Ipeirotis, Panos (2010) Analyzing the Amazon Mechanical Turk Marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2), p. 16-21.

Ipeirotis, P. (2009) Turker Demographics vs. Internet Demographics, *A Computer Scientist in a Business School*.

Johnson, J.A. (2005) Ascertaining the Validity of Individual Protocols from Web-Based Personality Inventories, *Journal of Research in Personality* 39(1), p. 103-129.

Mason, W. & Watts, D. (2009) Financial Incentives and the Performance of Crowds, *ACM SigKDD Explorations Newsletter*, 11(2), p. 100-108.

Meade, Adam W. & Craig, S. Bartholomew (2011) Identifying Careless Responses in Survey Data. Paper presented at the 26th Annual Meeting of the Society for Industrial and Organizational Psychology (April), Chicago, IL.

Oppenheimer, D.M., Meyvis, T. & Davidenko, N. (2009) Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power, *Journal of Experimental Social Psychology* 45, p. 867-872.

Paolacci, G., Chandler, J. & Ipeirotis, P.G. (2010) "Running Experiments on Amazon Mechanical Turk," *Judgment and Decision Making* 5, p. 411-419.

Peer, Eyal, Vosgerau, Joachim & Acquisti, Alessandro (2014) Reputation as a Sufficient Condition for Data Quality on Amazon Mechanical Turk, *Behavior Research Methods* 46(4), p. 1023-1031.

Rand, D.G. (2011) The Promise of Mechanical Turk: How Online Labor Markets Can Help Theorists Run Behavioral Experiments, *Journal of Theoretical Biology*, DOI: 10.1016/j.jtbi.2011.03.004.

Requester Best Practices Guide, *Amazon Mechanical Turk*, Amazon Web Services, 2011.

Ross, J., Irani, L., Silberman, M.S., Zaldivar, A. & Tomlinson, B. (2010) Who are the Crowdworkers?: Shifting Demographics in Mechanical Turk. In CHI EA '10: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems, p. 2863-2872, New York, NY, USA. ACM.

Suri, S. & Watts, D.J. (2011) Cooperation and Contagion in Web-Based, Networked Public Goods Experiments, *PLoS One*, 6(3), e16836.

Keywords: *MTurk, online data collection, crowdsourcing, attention check, Survey Monkey*

Relevance to Marketing Educators, Researchers and Practitioners: This presentation is designed to provide marketing researchers and practitioners with useful information on how to effectively use MTurk for online survey data collection while addressing common concerns and discussing key implementation decisions.

Author Information: L. Jean Harrison-Walker, Ph.D., a Professor of Marketing at the University of Houston-Clear Lake, is published in the Journal of Service Research, Journal of Services Marketing, Journal of Marketing for Higher Education, Journal of Marketing Theory and Practice, and Journal of Business-to-Business Marketing, as well as other journals and proceedings.

TRACK: Consumer Behavior and Marketing Research