

Reference: Curtis, M.G., Tran, T., Ibarra, M., Maxim, B. R., and Barchard, K.A. (2015, April). *MTurk Studies are quick and easy but not generalizable*. Paper presented at Western Psychological Association convention, Las Vegas, NV.

Contact Information: Kimberly A. Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, P.O. Box 455030, Las Vegas, NV, 89154-5030, USA, kim.barchard@unlv.edu

Abstract

Amazon Mechanical Turk (MTurk) is a crowd-sourcing website that allows individuals to contact others to complete tasks for compensation. Recently, social scientists have been using MTurk to advertise research studies. One advantage of using MTurk to contact potential participants is that researchers can obtain large samples quickly and easily. However, the diversity of MTurk users may be limited. The majority of users on MTurk reside in the United States and India (Paolacci & Chandler, 2014). Even within those two countries, MTurk users may not represent the general populations. The purpose of this study was to examine the demographic characteristics of MTurk users in the United States and India to determine if MTurk users are representative of the general populations in those countries.

Participants completed multiple-choice items measuring the following variables: sex, age, race, first language, the number of years that they have spoken English, and the country that they live in. Participants were excluded if their computer had an IP address that did not originate from India or the United States, or if they indicated that they did not live in India or the United States. This resulted in a sample of 733 participants (503 from India and 230 from the United States).

In both countries, there were slightly more men than women and participants averaged close to 30 years old. However, there were large differences in ethnicity and languages. In the US, most participants identified themselves as white and native English speakers. In India, most participants identified themselves as Asian and a variety of languages were spoken, the most common of which were Tamil, Malayalam, English, and Hindi.

MTurk users were not representative of the populations in either country. In the United States, African-Americans were extremely underrepresented. Approximately 13.2% of Americans are African-American (U.S. Census Bureau, 2014), but in our MTurk sample, only 8.3% were African-American. Similarly, the most common language in India is Hindi (Central Intelligence Agency, 2014), but only 10.1% of our sample spoke Hindi. Instead, the most common language in our MTurk sample was Tamil (40.8%), which is the 5th most popular language in India (Central Intelligence Agency, 2014).

Introduction

Technological advances have changed the way researchers collect data. One new option is to use Amazon Mechanical Turk (MTurk). Established in 2005, MTurk is a crowd-sourcing website that allows individuals to contract other individuals to complete tasks that computers cannot perform. Examples of such tasks include transcribing audio, manually updating information, and completing psychological studies. Frequently, this website is used by researchers to create and advertise studies. Hundreds of research studies have been conducted through MTurk and 16 of the top 30 universities in the United States collect data using MTurk in the areas of social psychology, cognitive psychology, linguistics, and other social sciences (Goodman, Cryder, & Cheema, 2013). MTurk provides researchers with the tools to design and implement studies, including the ability to set demographic limits on those who can participate, and the ability to offer compensation (Johnson & Borden, 2012). Using a searchable database, participants have the freedom to choose which tasks they would like to complete (Paolacci & Chandler, 2014).

MTurk includes 500,000 people from 190 countries who are mostly educated, unemployed, and of low socio-economic status (Paolacci & Chandler, 2014). However, MTurk samples have been shown to be significantly more diverse than typical Internet samples (Buhrmester, Kwang, & Gosling, 2011). The majority of participants reside in the United States and India and are within the age range of 18 to 30 years (Paolacci & Chandler, 2014). The MTurk subject pool is comprised of mostly female participants from the United States, whereas in India most of the participants are male (Antin & Shaw, 2012).

Using the Internet to collect data is relatively new. Because of this, the quality of experimental results that are based upon Internet samples have been frequently called into question. Various studies have concluded that data collected through MTurk is just as reliable as data obtained through traditional studies (Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett, 2013). MTurk has also been shown to have a broader demographic subject pool than

comparable Internet samples and to be significantly more diverse than the standard American college populations used in many studies (Buhrmester, Kwang, & Gosling, 2011). Although MTurk data are just as valid as the data from traditional studies, the particular participants who use MTurk could potentially sway the outcome of a study (Steelman, Hammer, & Limayem, 2014). For example, the inclusion of non-US participants can create statistically significant differences in study results, because of this, researchers must be wary of making broad generalizations using such data (Steelman, Hammer, & Limayem, 2014). Even if researchers restrict their participants to people who are from a single country, it is possible that MTurk samples do not represent the general population of that country. The purpose of this study is to examine the demographic characteristics of MTurk samples in the United States and India (the countries with the largest number of MTurk users), to determine if MTurk users are representative of the U.S. and India populations and if results can be generalized to the general populations in each country.

Table 1
Percent of Participants with Various Demographic Characteristics

| <u>Variable</u> | <u>Level</u> | <u>Country</u> | | |
|-----------------------|---|----------------------|--------------|-----|
| | | <u>United States</u> | <u>India</u> | |
| Sex | Male | 57.8 | 59.8 | |
| | Female | 42.2 | 40.0 | |
| Race | American Indian or Alaska Native | 2.2 | 2.2 | |
| | Asian | 9.6 | 90.1 | |
| | Black or African American | 8.3 | 0.2 | |
| | Native Hawaiian or Other Pacific Islander | 0.4 | | |
| | White | 80.4 | 1.6 | |
| | Other | 2.6 | 4.6 | |
| | The numbers of years participants have spoken English | Bengali | 0.4 | 1.4 |
| | | Cantonese | 0.9 | |
| Cebuano | | 0.4 | | |
| English | | 95.7 | 15.7 | |
| Farsi | | | 0.2 | |
| Guararati | | | 1.0 | |
| Hindi | | | 10.1 | |
| Kannada | | | 1.8 | |
| Korean | | 0.4 | | |
| Malayalam | | | 17.1 | |
| Marathi | | | 2.0 | |
| Oriya | | | 1.2 | |
| Punjabi | | | 0.6 | |
| Spanish | | 1.3 | | |
| Tagalog | | 0.4 | | |
| Tamil | 0.4 | 40.8 | | |
| Telugu | | 6.4 | | |
| Urdu | | 1.0 | | |
| Vietnamese | | 0.2 | | |
| Other, please specify | | 0.6 | | |

Method

Participants

A total of 733 participants completed this study. Of these, 503 lived in India and 230 lived in the United States. We deliberately collected more data from India than the United States, so that the data could be used for another study. This study cannot comment on the relative numbers of people from the United States and India who use MTurk.

Measures

Demographics Survey

The demographics survey consists of multiple-choice items measuring the following variables: sex, age, race, first language, the number of years that they have spoken English, and the country that they live in.

Procedure

The demographics survey was administered as part of a larger study that took about 15 minutes to complete. The advertisement was shown on MTurk. Participants clicked on a link that took them to a Qualtrics survey. When they submitted the Qualtrics survey, they were automatically paid 35 cents through MTurk.

Results

In the United States and India, there were more men than women (United States: 57.8%, India, 59.8%). In both countries, participants were older than college samples (United States: mean age 30.9, India: mean age 30.5), and there was a wide range of ages (United States: SD 9.9 years, India: SD 8.7

years). There were large differences in languages and race between the two countries, as would be expected (see Table 1). In the United States, most participants identified themselves as white and native English speakers. In India, most

participants identified themselves as Asian. A variety of languages were spoken, the most common of which were Tamil, Malayalam, English, and Hindi.

Discussion

The use of websites such as MTurk is a relatively new way of collecting psychological data. MTurk samples have been shown to yield results that are just as valid as those samples obtained through more traditional means such as phone or mail surveys (Steelman, Hammer, & Limayem, 2014). This new way of collecting data has made it easier for researchers to collect large data sets in a short amount of time. However, researchers should also consider the limitations of MTurk samples.

Our study showed that MTurk participants are not representative of the general population in the United States. In particular, African-Americans are underrepresented. The United States census shows that 13.2% of Americans are African-American (U.S. Census Bureau, 2014), but in our MTurk sample, only 8.3% of users were African-American. Other minorities were over-represented in our MTurk sample. For example, according to the census, Asian Americans comprise 5.3% of the population in the United States (U.S. Census Bureau, 2014), yet in our study 9.3% of participants from the United States were Asian.

Similarly, MTurk users in India were not representative of the entire Indian population. Although Hindi is the most widely spoken language throughout the country with 41% of the overall population speaking the language (Central Intelligence Agency, 2014), Tamil was the most likely to be listed as the first language of our participants from India: 40.8% of users from India said this was their first language. Tamil only ranks as the 5th most popular language in India with 5.9% of the overall population speaking the language (Central Intelligence Agency, 2014). Moreover, English (15.7%) was also a very common first language among the participants from India because English is an official language of India that is spoken for national, political, and commercial communication (Central Intelligence Agency, 2014).

When examining the Indian languages prevalent in our study in more detail, it seems that the participants from India may have come primarily from particular Indian states. The native languages that dominated our sample are derived from the Dravidian languages that originate from the southern portion of the country (see Figure 1; Schwartzberg, 1980). Of these languages, the most prevalent in our study were Tamil (40.8%), Malayalam (17.1%), and Telugu (6.4%). Thus, northern India may have been under-represented in our MTurk sample.

Researchers should be cautious about making broad generalization from samples acquired through MTurk. The data in our study shows that there may be large discrepancies between the demographic characteristics of MTurk samples in the United States and India. There may have also been discrepancies within the respective countries as wholes, especially in terms of ethnicity and language. To create more representative samples, researchers could collect data on MTurk using stratified sampling or they could collect data within narrower, more homogeneous groups and forsake generalization to wider groups.

Future research should be dedicated specifically to exploring the characteristics of MTurk users who are likely to complete psychological studies. These analyses were conducted as a secondary study of a larger dataset; thus, the variables that were available to us were limited. In particular, one of the limitations of our study was the way in which ethnicities were classified. Ethnic classifications are different in the United States and India. India does not recognize racial or ethnic groups but does recognize certain castes and tribes (The Register General & Census Commissioner, India, 2011). This limited the ways participants in India could identify their ethnicity: they were only able to cite themselves as

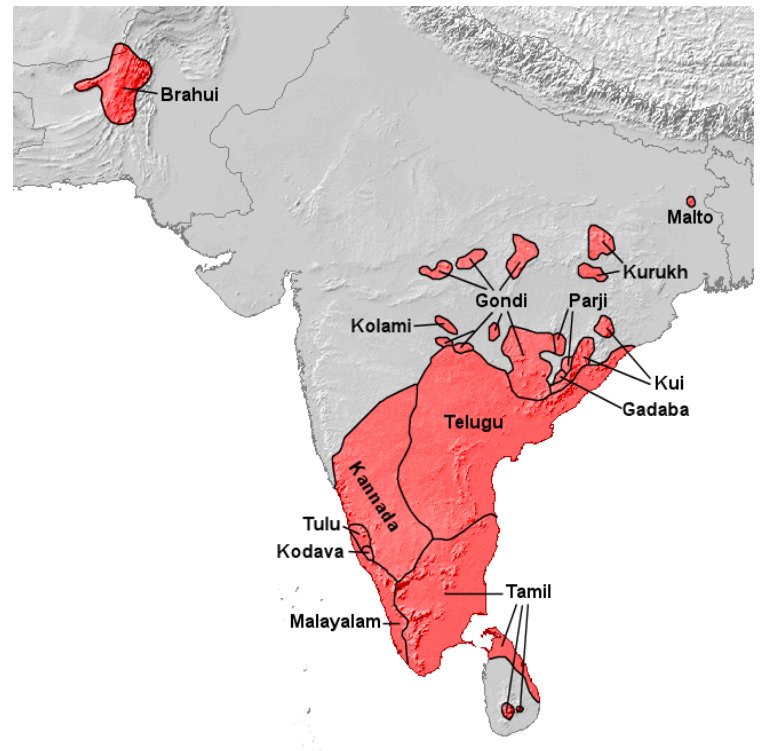


Figure 1
Distribution of India languages as of 1992 (Schwartzberg, 1992)

being Asian on our survey. They were unable to adequately identify themselves with a specific tribe. Therefore, future studies should be tailored to the specific classifications used by each country.

Another limitation of our study was that we did not measure certain core demographic factors, such as the socioeconomic status and education level. Future research should include these factors to provide a clearer picture of who uses MTurk and whether they are representative of the general populations in those countries.

Finally, future research should use larger samples. When trying to estimate the proportion of MTurk users who belong to minority groups (such as ethnic or language groups), it is essential to have very large samples if these proportions are to be estimated accurately. In particular, the United States sample, with only 230 participants, was somewhat small for making conclusions about the proportion of people who belong to each ethnic group. Ideally, future research would be based upon the entire population of MTurk users, at least for those variables that all users complete in their basic user profile (e.g., country, sex, and age).

In conclusion, while MTurk is a fast and efficient way of collecting data, researchers should be cautious about generalizing findings from MTurk samples to the general population because not all demographic groups are adequately represented.

References

- Antin, J., & Shaw, A. (2012). Social desirability bias and self-reports of motivation: a study of amazon mechanical turk in the US and India. *Proceedings of the SIGCHI Conference on human factors in computing systems*, 2925-2934.
- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351-368. doi:10.1093/pan/mpr057
- 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
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29(6), 2156-2160. doi:10.1016/j.chb.2013.05.009
- Central Intelligence Agency. (2014). India. *In The World Factbook*. Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/geos/in.html>
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46(1), 112-130. doi:10.3758/s13428-013-0365-7
- 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(3), 213-224. doi:10.1002/bdm.1753
- National Portal of India. (21, August, 2014). *India at a Glance*. Retrieved from <http://india.gov.in/india-glance/profile>.
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184-188. doi:10.1177/0963721414531598
- The Register General & Census Commissioner, India. (2011). Scheduled Castes & Scheduled Tribes Population. Retrieved from [http://www.censusindia.gov.in/\(Sj11tumil0zmzt155dpqscx45\)\)/Census_Data_2001/India_at_Glance/scst.aspx](http://www.censusindia.gov.in/(Sj11tumil0zmzt155dpqscx45))/Census_Data_2001/India_at_Glance/scst.aspx)
- Schwartzberg, J. E. (1980). A Historical Atlas of South Asia. *American Anthropologist*, 82(1), 143-144. doi: 10.1525/aa.1980.82.1.02a00280
- Schwartzberg, J. E. (1992). *Language families and branches, languages and dialects*. New York, NY: Oxford University Press.
- Shapiro, D. N., Chandler, J. J., & Mueller, P. A. (2013). Using mechanical turk to study clinical and subclinical populations. *Clinical Psychology Science*, 1(2), 213-220. doi:10.1177/2167702612469015
- Steelman, Z. R., Hammer, B. I., & Limayem, M. (2014). Data collection in the digital collection in the digital age: innovative alternative to student samples. *MIS Quarterly*, 38(2), 355-A20.
- Suri, S., & Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS One*, 6(3), e16836. doi:10.1371/journal.pone.0016836
- U.S. Census Bureau. (2014, July 8). *State & county quickfacts*. Retrieved from <http://quickfacts.census.gov/qfd/states/00000.html>.