

## **An Analysis of the Use of Amazon's Mechanical Turk for Survey Research in the Cloud**

Marc Dupuis<sup>1</sup>, Barbara Endicott-Popovsky<sup>2</sup> and Robert Crossler<sup>3</sup>

<sup>1</sup>Institute of Technology, University of Washington, Tacoma, United States of America

<sup>2</sup>The Information School, University of Washington, Seattle, United States of America

<sup>3</sup>College of Business, Mississippi State University, Starkville, United States of America

[marcjd@uw.edu](mailto:marcjd@uw.edu)

[endicott@uw.edu](mailto:endicott@uw.edu)

[rob.crossler@msstate.edu](mailto:rob.crossler@msstate.edu)

**Abstract:** Survey research has been an important tool for information systems researchers. As technologies have evolved and changed the manner in which surveys are administered, so have the techniques employed by researchers to recruit participants. Crowdsourcing has become a common technique to recruit participants for different kinds of research, including survey research.

This paper examines the role of one such crowdsourcing platform, Amazon's Mechanical Turk (MTurk). MTurk allows everyday people to create an account and become a worker to perform various tasks, called HITS (human intelligence tasks). HITS are posted by requestors, which may be researchers, corporations, or other entities that have generally simple tasks that can be performed through crowdsourcing.

We examine MTurk in the context of five different surveys conducted using the MTurk platform for the recruitment phase of the studies. Our discussion includes both practical things to consider when using the platform and an analysis of some of our findings. In particular, we explore the use of qualifiers in conducting studies, things to consider in both longitudinal and cross-cultural studies, the demographics of the MTurk population, and ways to control and measure quality.

Although MTurk participants do not perfectly represent the U.S. population from a demographics standpoint, we found that they provide good overall diversity on several key indicators. Furthermore, this diversity that MTurk samples provide will often be as good, if not better, than the typical participants recruited for research (e.g., college sophomores).

Similar to most recruitment methods, using MTurk to conduct research does have its drawbacks. Nonetheless, the evidence does not indicate that these drawbacks are either significant enough to preclude the use of such a platform, or in any way more significant than the drawbacks associated with other techniques. In fact, quality is generally high, the cost is low, and the turnaround time is minimal. We do not suggest that MTurk should replace other techniques for participant recruitment, but rather that it deserves to be part of the discussion.

**Keywords:** crowdsourcing, surveys, Mechanical Turk, cloud

### **1. Introduction**

Survey research has a long and rich history that can be traced back to the censuses of the Old Kingdom (3<sup>rd</sup> millennium BC) in ancient Egypt (Janssen 1978). Administration of surveys during this time period occurred in-person, but new capabilities and technologies that emerged in the twentieth century led to additional administration techniques. For example, the use of mail and telephones to administer surveys became the norm (Armstrong and Overton 1977; Fricker et al. 2005; Kaplowitz et al. 2004; Kempf and Remington 2007). Toward the end of the century, we would again see the use of new administration techniques based on emerging technologies. The Internet allowed for the use of email to collect survey data, followed shortly thereafter with web-based administration of surveys (Krathwohl 2004; Schutt 2012; Sheehan 2001).

Along with these shifts in administration techniques, there have also been new methods employed to recruit participants within any one of these techniques. Most recently, this has included crowdsourcing to recruit participants to complete surveys on the Web (Howe 2006; Kittur et al. 2008; Mahmoud et al. 2012; Ross et al. 2010). This paper examines the use of a particular crowdsourcing platform to perform this type of research, Amazon's Mechanical Turk (MTurk). In particular, we will begin by discussing some of the background literature. This is followed by an examination of using the MTurk platform in practice, along with some analysis from several studies. Finally, some concluding thoughts will be given on the use of MTurk to conduct research in the cloud.

The paper makes an important contribution by further exploring the role the MTurk platform may play in research. It expands on earlier research in this area by providing an up-to-date analysis of current trends, demographics, and uses of MTurk as a tool for researchers. Survey research has been an incredibly important tool for researchers, including within the information systems domain (Anderson and Agarwal 2010; Atzori et al. 2010; Burke 2002; Chen and Kotz 2000; Crossler 2010; LaRose et al. 2008; Liang and Xue 2010; Zeng et al. 2009). Thus, it is an important tool to examine in further in the context of cloud security research.

## **2. Background literature**

Traditionally, individuals and organizations interested in having work performed in exchange for compensation were limited by the relatively small marketplace available to them. Methods that could be employed to expand the available marketplace were often expensive, time-consuming, and not always practical. However, the Internet and its broad spectrum of users have changed this dynamic considerably. It has made the expansion of the marketplace often cost effective, quick, and practical. In this section, we discuss crowdsourcing in general and the use of Amazon's Mechanical Turk (MTurk) in particular.

### **2.1 Crowdsourcing**

According to Mason and Watts (2010), in crowdsourcing "potentially large jobs are broken into many small tasks that are then outsourced directly to individual workers via public solicitation" (p. 100). Crowdsourcing has become quite popular in recent years due primarily to the possibilities the Internet provides for individuals and organizations. An early example of the power of crowdsourcing is iStockphoto. Stock photos that used to cost hundreds of dollars to license from professionals could often be had for no more than a dollar a piece (Howe 2006). Rather than just professionals contributing images to the site, students, homemakers, and other amateurs would contribute several images to earn some extra money. The condition that allows crowdsourcing to work so effectively is that many of the *workers* perform tasks during their spare time. In other words, it is generally not their main source of income, but rather supplements other possible income sources.

These workers are not limited to a few dollars at a time either. Since 2001, corporate R&D departments have been using Eli Lilly's InnoCentive to find intellectual talent that can solve complex problems that have been stumping their own people for a while (Howe 2006). These solvers, as they are called, may earn anywhere from \$10,000 to \$100,000 per solution. While these types of workers may be limited to those with the requisite skills and talent, MTurk is available to the masses.

### **2.2 Amazon's Mechanical Turk**

MTurk allows everyday people to create an account and perform various tasks as workers. These tasks are called HITs (human intelligence tasks) and are posted by requesters (Mason and Watts 2010). HITs usually pay anywhere from \$0.01 to a few dollars each, generally depending on the skill level required and the amount of effort involved. The opportunities for researchers are great, but there are naturally several questions that arise, such as: demographics, quality, and cost.

#### **2.2.1 Demographics**

First, the composition of any subject pool is always of great interest to the researcher. The MTurk workers do represent a special segment of the population; namely, those that have Internet access and are willing to complete HITs for minimal pay. However, this is true of any participants that agree to participate in social science research (Horton et al. 2011).

MTurk participants are also generally younger than the population they are meant to represent, although their age is generally more representative than what may often be found in university subject pools (Paolacci et al. 2010). Additionally, U.S. workers are disproportionately female, while workers from India are disproportionately male (Horton et al. 2011; Ipeirotis et al. 2010; Paolacci et al. 2010).

Nonetheless, they are generally comparable to other populations often recruited for research, including Internet message boards (Paolacci et al. 2010). Beyond age and gender, MTurk participants also represent a diverse range of income levels (Mason and Watts, 2010). In a study comparing MTurk participants with other Internet samples, Buhrmester, Kwang, and Gosling (2011) found that "MTurk participants were more demographically diverse than standard Internet samples and significantly more diverse than typical American college samples" (p. 4). Thus, the demographics of

MTurk participants are comparable to other types of samples often used, and in some instances they may be superior.

### *2.2.2 Quality*

If demographics are not an issue in using MTurk participants, what about the quality of the results? Interestingly, quality has not been a major limiting factor in using MTurk for research purposes. In some instances, quality may be better than university subject pools and those recruited from Internet message boards. One way to measure quality is to devise one or more test questions, also called “catch trials”. These types of questions have obvious answers that anyone paying adequate attention to the wording should be able to answer correctly with ease. In one study, MTurk participants had a failure rate on the catch trials of 4.17 percent, compared to 6.47 and 5.26 percent for university subject pool participants and Internet message boards participants, respectively (Paolacci et al., 2010, p. 416). A majority of incorrect answers can generally be associated with only a small subset of participants rather than widespread gaming of the system (Kittur et al., 2008).

Likewise, their survey completion rate for MTurk participants (91.6 percent) was significantly higher than those recruited from Internet message boards (69.3 percent) and not that far behind university subject pool participants (98.6 percent). Additionally, quality is not impacted by either the amount paid for a HIT or the length of time required to complete the task, although both of these factors will impact how long it takes to recruit a given number of participants (Buhrmester et al., 2011; Mason and Watts, 2010).

Finally, another important test for quality is the psychometric properties of completed surveys. Similar to the other measures for quality, MTurk participants had absolute mean alpha levels in the good to excellent range (Buhrmester et al., 2011, pp. 4–5). This was true for all compensation levels and for all scales. Test-retest reliability was also very high with a mean correlation of 0.88. All of this is comparable to traditional methods and suggests that the psychometric properties of survey research from MTurk participants are acceptable for academic research purposes.

### *2.2.3 Cost*

The final question that we will address from the background literature relates to cost. Is MTurk a cost-effective solution for conducting academic research? Multiple studies have demonstrated that the cost to use MTurk is not only reasonable, but quite low (Buhrmester et al., 2011; Horton et al., 2011; Kittur et al., 2008; Mason and Watts, 2010; Paolacci et al., 2010). In one instance, a HIT that paid only \$.01 to answer two questions received 500 responses in 33 hours (Buhrmester et al., 2011). In another instance, MTurk participants were compared with offline lab participants. Whereas offline lab participants received a \$5 show-up fee, their MTurk counterparts received only \$0.50 (Horton et al., 2011).

Thus, MTurk provides an opportunity for researchers to perform research on the Web at often a fraction of the cost of traditional methods. Quality and demographics are comparable to these other methods, while the speed at which data can be collected is generally superior. In the next section, we will continue this discussion with an examination of results from recent studies we conducted using the MTurk platform.

## **3. MTurk in practice: considerations and analysis**

There are many things to consider when using MTurk. This section is not meant to be exhaustive, but rather will touch on a few important considerations while using the MTurk platform. Additionally, the MTurk platform offers a powerful API that provides additional capabilities. These capabilities often go beyond what can be done through the Web platform. However, for many researchers this may not prove practical if one does not have the requisite technical skills. Therefore, the focus here will be on what can be done using the Web interface on MTurk, possible workarounds, and inherent limitations. This will be based on multiple HITs the authors have posted over the past 12 months.

### **3.1 Recruitment**

Using MTurk to recruit participants is relatively straightforward. After an account is created, you must fund the account with an amount adequate to cover both the compensation amount you plan on providing to participants and Amazon’s fee, which is 10 percent of the total amount paid to the workers (Amazon, 2013). For example, if the goal is to recruit 300 participants at \$0.50 each, then the

account should be pre-funded \$165. It is a good idea to build some flexibility into how much you fund your account so that you can easily and quickly adjust the amount paid per HIT, if necessary.

Another important consideration is on the description provided for the project. The description should be clear, accurate, short, and simple (Amazon Web Services LLC, 2011). Furthermore, if a survey is being conducted then it should be emphasized how short the survey is or how quick it will be to complete it. This is important given that MTurk provides a marketplace in which the primary goal is for each worker to maximize total income as a product of his or her time. The use of ambiguous terms (e.g., few) is less likely to be effective than explicit descriptions (e.g., 5 multiple-choice questions).

Once the requestor has created a project, a batch must be created for that project before workers are able to view it. Batches may be extended longer or ended early.

### 3.2 Compensation and time

In a qualification survey we conducted, we found that price was a large factor in how quickly assignments were completed. After it was determined that the rate was too low, the amount paid for the HIT was increased. This was done multiple times with each subsequent HIT available only to those that did not complete one of the earlier ones. Below is a table that illustrates the amount paid, total time available, average time per assignment, and number of completed HITs.

**Table 1:** Qualifying survey HIT results

HIT #	Compensation	Time Available	Average Time Per Assignment	Completed
1	\$0.05	≈ 2 hours	1 minute, 29 seconds	10
2	\$0.11	≈ 1 day, 9 hours	2 minutes, 5 seconds	368
3	\$0.12	≈ 1 day	2 minutes, 9 seconds	106
4	\$0.13	≈ 19 hours	2 minutes, 25 seconds	200

A couple of quick observations are worth noting. First, the average time per assignment increased as the compensation increased ( $r = 0.985$ ;  $p < .05$ ). Second, it appears that some workers may have a specific price point that must be met prior to completing a HIT. The number of completed assignments at each price point was expected to be higher, but the wording of the HIT may have played a role in not attracting a greater number of workers.

In another HIT administered August of 2012 to U.S. residents, the worker was required to complete a relatively long survey that was estimated to take between 15 and 20 minutes. Workers were paid \$0.75 for completing the HIT. Results from 303 workers were obtained in approximately 12 hours with an average time per assignment of 10 minutes and 39 seconds. This suggested that the amount paid could be lower.

In July of 2013, a HIT for a survey that was similar in length and also limited to U.S. residents was created. Workers were paid \$0.50 for this HIT and it was completed in approximately five hours. The average time per assignment was nine minutes and 28 seconds. An identical HIT was created for residents of India with an average time per assignment of nine minutes and 27 seconds, virtually identical to the U.S. population.

### 3.3 Qualifications

When creating the project, you may also specify certain predetermined qualifiers, which allows the requestor to limit the HITs only to those meeting the qualifications, such as location (e.g., United States). The location qualifier allows you to choose the country, but does not provide any additional granularity (e.g., state). Other built-in qualifiers include the number of HITs completed and the acceptance rate.

In addition to predetermined qualifiers, requestors (i.e., the researcher) may also create certain qualifications and assign scores between 0 and 100. You can have workers complete a short survey through the MTurk platform or an external survey platform (e.g., Qualtrics, Survey Gizmo, Survey Monkey) and update worker qualifications based on this data. The most efficient method to do this using the Web-based MTurk platform involves creating the qualifications and then downloading the worker CSV (comma separated values) file that contains historical information on all of your workers. The CSV file contains requestor-created qualifications, columns with current values for requestor-

created qualifications, as well as columns to update the qualifications. The requestor can then update the CSV file and upload it back to the system, which Amazon will process and update accordingly. Qualifications can be assigned manually through the Web-based platform, but for larger numbers of workers for which one would like to assign qualifications to, this becomes quite impractical.

### 3.4 Quality

Quality is an important concern for any researcher. The primary test for quality that will be analyzed here involves the use of quality control questions, also referred to as catch trials. Five studies were conducted that included a quality control question. Studies four and five were follow-up surveys to studies two and three and only those that passed the quality control question in the earlier survey were eligible to complete the second survey, which was administered approximately five weeks later. The table below illustrates these results.

**Table 2:** Quality control question failure rate

Study	Population	Total Number of Submissions	Failure Rate
1	U.S.	303	2.31%
2	U.S.	170	8.82%
3	India	212	29.25%
4	U.S.	110	2.73%
5	India	131	15.27%

A couple of observations are worth noting. First, the failure rate for residents from India is very high in general and in comparison to residents from the U.S. ( $r = 0.899$ ;  $p < .05$ ). It is unclear why the failure rate is so much higher. Possibly, language or cultural factors may play a role.

Second, while the failure rate for U.S. residents is still relatively low, it is unclear why this rate jumped considerably from study one, conducted in August of 2012, to the second study. The quality control questions were slightly different and the HIT paid a different amount. It is unclear if either of these factors contributed to the difference in failure rates.

### 3.5 Longitudinal studies

Conducting longitudinal studies is an important method for many different types of research. The discussion earlier on assigning qualifications to workers is necessary for fixed-sample longitudinal studies. Basically, those that successfully complete the first phase of the study are given a qualification for future phases. Then, when creating subsequent projects on the MTurk platform, the requestor will limit the HIT to only those with the requisite qualification.

This was done for a longitudinal study examining the Edward Snowden situation. However, workers will not necessarily see the HIT as anything special. In other words, they will generally be as likely to view the HIT as any other HIT. In our follow-up survey, we increased the pay from \$0.50 to \$0.65. Workers from India responded quite well, which may be due in part to the relatively high value the HIT pays as well as possibly fewer HITs available overall to residents of India. Approximately 85 workers from India completed the second survey within 48 hours compared to only 18 from the U.S.

In a comparative cross-cultural longitudinal study, these numbers become quite problematic. The API and associated tools do provide a mechanism to send an email to workers. However, this was not as simple using the Web-based MTurk platform. Email messages could be sent through the platform, but involved navigating to the project that contained the results from phase one, filtering out those that did not pass the quality control question, and finally emailing only those that had not yet completed phase two.

Nonetheless, completing this process helped immensely. Messages were sent to the appropriate workers informing them of the HIT with the URL. The number of U.S. participants increased from approximately 18 to over 90 within 48 hours. Likewise, the number of participants from India increased from approximately 85 to 120. A second message was sent to both groups of workers, which resulted in a total of 110 submissions from the U.S. and 131 from India. Therefore, it is possible to conduct longitudinal research using the MTurk platform, but extra steps may be required to obtain

an acceptable number of responses. Other longitudinal designs that do not require a fixed-sample would alleviate some of these issues, but would also introduce new ones in the process.

### **3.6 International populations**

Throughout the discussion, we have examined specific facets of using MTurk as a research platform. Within this discussion, there has been some comparison between participants from the U.S. and those from India. In summary, participants from India have a significantly higher failure rate on quality control questions, but are also more likely to participate in subsequent phases of a study when a fixed-sample longitudinal design is employed.

Originally, workers from other countries were sought in addition to the U.S. and India to participate in the Snowden study. However, it became untenable when only one worker from the U.K. completed the HIT for the first phase of the study over a 16 hour time-period. The amount paid for the HIT was increased from \$0.50 to \$0.75, but no additional workers participated. Likewise, there was only one worker from Sweden that responded to the survey over a 50-hour time-period.

Thus, consistent with other research that has used MTurk, most workers are from the U.S. and India (Horton et al. 2011; Ipeirotis et al. 2010). While the studies examined here did not rule out other countries (e.g., Canada), the evidence at least suggests that requestors will be the most successful when recruiting workers from either India or the U.S.

### **3.7 Demographics**

Earlier, we discussed the demographics of the MTurk workers based on prior research. In this section, we will revisit the issue of demographics and examine the composition of MTurk workers from four different studies we conducted with a more in-depth analysis of workers that completed the qualifying survey.

In the table that follows, we look at the age, gender, and educational attainment levels of MTurk workers from four different studies. What may be the most interesting are the changes in gender distribution between males and females. Other research has generally found a greater percentage of female workers than males (Horton et al. 2011; Ipeirotis et al. 2010; Paolacci et al. 2010), but this is not found in studies four and five. It may be due to the limited time the HIT was available. In other words, females may be more apt to complete a HIT during certain hours of the day, while males may be more apt to do so during other hours of the day. The qualifying survey shows that approximately 18.4 percent of MTurk workers are unemployed, retired, or a homemaker and 10.3 percent are students.

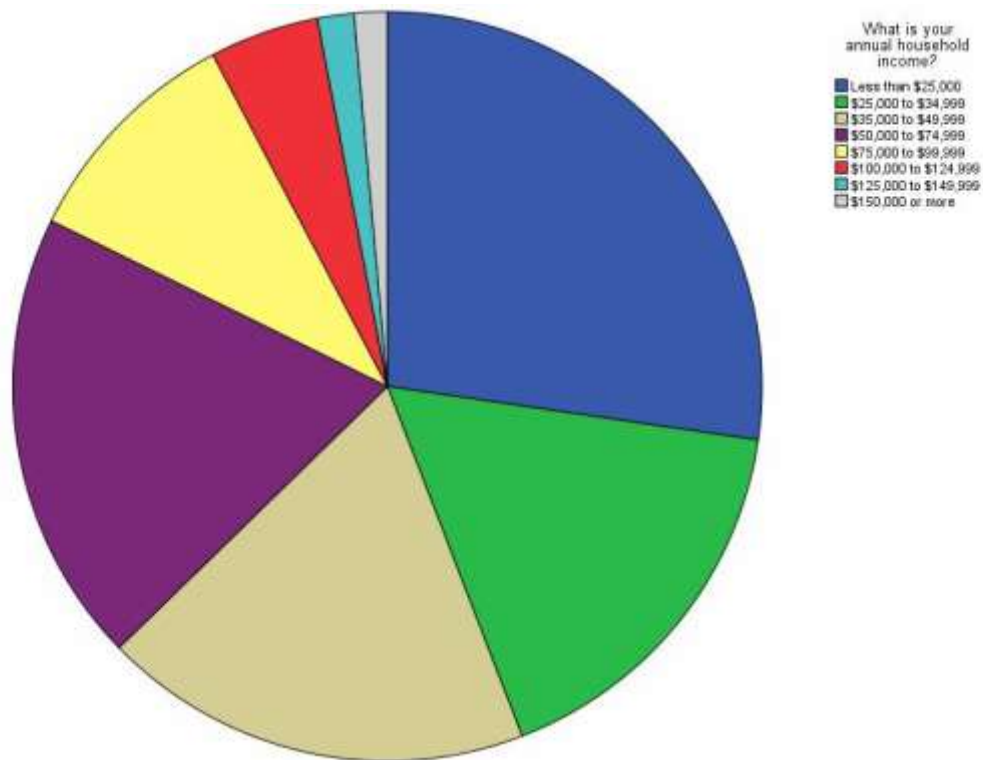
The study that would largely mitigate this possible effect is the qualifying survey due to the length of time it was available and the overall sample size. While the gender distribution is not heavily male as it is in studies four and five, it is also not heavily female as it has been in other studies and as is reflected in study one. In fact, the ratio is almost 50:50 and largely consistent with the U.S. population as a whole.

**Table 3: Age, gender, and educational attainment levels**

	<b>U.S. Sample, Study 1</b>	<b>U.S. Sample, Study 4</b>	<b>India Sample, Study 5</b>	<b>Qualifying Survey</b>	<b>U.S. Population<sup>1</sup></b>
<b>Sample Size</b>	296	101	107	702	--
<b>Time Period</b>	August 2012	August 2013	August 2013	July 2013	2011/2012
<b>Age</b>					
18-29	41.55%	37.62%	41.12%	48.4%	22.0%
30-39	25.34%	30.69%	40.19%	25.7%	17.0%
40-49	16.89%	14.85%	8.41%	12.2%	18.2%
50-59	11.49%	9.90%	4.67%	10%	18.1%
60+	4.73%	6.93%	5.61%	3.6%	24.7%
<b>Gender</b>					
Male	42.6%	63.37%	64.49%	50.4%	49.1%
Female	57.1%	36.63%	35.51%	49.3%	50.9%
<b>Education</b>					
Some High School	N/A	0.99%	.93%	1.1%	8.58%
High School (or GED)	N/A	8.91%	.93%	11.1%	30.01%
Some College	N/A	32.67%	9.35%	29.4%	19.46%
College Graduate	N/A	47.52%	57.01%	48.8%	27.59%
Master's / Professional Degree	N/A	6.93%	31.78%	7.7%	8.4%
Doctorate	N/A	2.97%	--	1.9%	1.36%

If we further examine the data from the qualifying survey, we can explore a few additional demographic indicators. For example, the pie chart that follows illustrates the diversity in household income levels from MTurk workers. As we can see, there is a broad spectrum of household income levels represented.

<sup>1</sup> Source: U.S. Census Bureau, 2011/2012



**Figure 1:** Annual household income distribution

Finally, we will examine the geographic distribution of MTurk workers. A table that compares the geographic distribution of the U.S. population with the MTurk workers that completed the qualifying survey follows.

**Table 4:** Regional distribution of U.S. MTurk workers

Region	Qualifying Survey	U.S. Population
Northeast	19.4%	18.01%
Midwest	24.7%	21.77%
South	32.2%	36.91%
West	23.6%	23.31%

The geographic distribution of MTurk workers in the U.S. closely resembles that of the U.S. population. To the extent the MTurk workers do not closely resemble the U.S. population, they do nonetheless provide a much greater degree of diversity on key demographic indicators than the typical college sophomore (Sears 1986).

#### 4. Conclusion

The MTurk platform is a relatively new method that can be used to recruit participants in not only survey research, but several other types of research activities as well. Similar to most recruitment methods, using MTurk to conduct research does have its drawbacks (e.g., incentives, motivation, quality) (Horton et al., 2011; Mason and Watts, 2010). Nonetheless, the evidence does not indicate that these drawbacks are either significant enough to preclude the use of such a platform, or in any way more significant than the drawbacks associated with other techniques (Buhrmester et al. 2011; Gosling et al. 2004; Sears 1986). In fact, quality is generally high, the cost is low, and the turnaround time is minimal. We do not suggest that MTurk should replace other techniques for participant recruitment, but rather that it deserves to be part of the discussion.



## References

- Amazon, 2013. Amazon Mechanical Turk: Requestor FAQs [WWW Document]. URL [https://requester.mturk.com/help/faq#how\\_fees\\_cost\\_computed](https://requester.mturk.com/help/faq#how_fees_cost_computed) (accessed 5.5.13).
- Amazon Web Services LLC, 2011. Requestor Best Practices Guide [WWW Document]. URL [http://mturkpublic.s3.amazonaws.com/docs/MTURK\\_BP.pdf](http://mturkpublic.s3.amazonaws.com/docs/MTURK_BP.pdf) (accessed 4.15.13).
- Anderson, C.L., Agarwal, R., 2010. Practicing Safe Computing: A Multimethod Empirical Examination of Home Computer User Security Behavioral Intentions. *MIS Q.* 34, 613–643.
- Armstrong, J., Overton, T., 1977. Estimating nonresponse bias in mail surveys. *J. Mark. Res.* 14, 396–402.
- Atzori, L., Iera, A., Morabito, G., 2010. The internet of things: A survey. *Comput. Networks* 54, 2787–2805.
- Buhrmester, M., Kwang, T., Gosling, S.D., 2011. Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data? *Perspect. Psychol. Sci.* 6, 3–5.
- Burke, R., 2002. Hybrid recommender systems: Survey and experiments. *User Model. User-Adapt. Interact.* 12, 331–370.
- Chen, G., Kotz, D., 2000. A survey of context-aware mobile computing research. Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College.
- Crossler, R.E., 2010. Protection Motivation Theory: Understanding Determinants to Backing Up Personal Data. *Koloa, Kauai, Hawaii*, p. 10.
- Fricker, S., Galesic, M., Tourangeau, R., Yan, T., 2005. An Experimental Comparison of Web and Telephone Surveys. *Public Opin. Q.* 69, 370–392.
- Gosling, S.D., Vazire, S., Srivastava, S., John, O.P., 2004. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *Am. Psychol.* 59, 93.
- Horton, J.J., Rand, D.G., Zeckhauser, R.J., 2011. The online laboratory: Conducting experiments in a real labor market. *Exp. Econ.* 14, 399–425.
- Howe, J., 2006. The Rise of Crowdsourcing. *Wired* 14.
- Ipeirotis, P.G., Provost, F., Wang, J., 2010. Quality management on Amazon Mechanical Turk, in: *Proceedings of the ACM SIGKDD Workshop on Human Computation*. ACM, Washington DC, pp. 64–67.
- Janssen, J., 1978. The Early State in Ancient Egypt, in: Claessen, H.J., Skalník, P. (Eds.), *The Early State*. Walter de Gruyter, pp. 213–234.
- Kaplowitz, M.D., Hadlock, T.D., Levine, R., 2004. A comparison of web and mail survey response rates. *Public Opin. Q.* 68, 94–101.
- Kempf, A.M., Remington, P.L., 2007. New Challenges for Telephone Survey Research in the Twenty-First Century. *Annu. Rev. Public Health* 28, 113–126.
- Kittur, A., Chi, E.H., Suh, B., 2008. Crowdsourcing user studies with Mechanical Turk, in: *Proceedings of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems*. ACM, Florence, Italy, pp. 453–456.
- Krathwohl, D., 2004. *Methods of educational and social science research : an integrated approach*, 2nd ed. ed. Waveland Press, Long Grove Ill.
- LaRose, R., Rifon, N.J., Enbody, R., 2008. Promoting Personal Responsibility for Internet Safety. *Commun. ACM* 51, 71–76.
- Liang, H., Xue, Y., 2010. Understanding Security Behaviors in Personal Computer Usage: A Threat Avoidance Perspective. *J. Assoc. Inf. Syst.* 11, 394–413.
- Mahmoud, M.M., Baltrusaitis, T., Robinson, P., 2012. Crowdsourcing in emotion studies across time and culture, in: *Proceedings of the ACM Multimedia 2012 Workshop on Crowdsourcing for Multimedia*. ACM, Nara, Japan, pp. 15–16.
- Mason, W., Watts, D.J., 2010. Financial incentives and the performance of crowds. *ACM SigKDD Explor. Newsl.* 11, 100–108.
- Paolacci, G., Chandler, J., Ipeirotis, P., 2010. Running experiments on amazon mechanical turk. *Judgm. Decis. Mak.* 5, 411–419.
- Ross, J., Irani, L., Silberman, M., Zaldivar, A., Tomlinson, B., 2010. Who are the crowdworkers?: shifting demographics in mechanical turk, in: *Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, pp. 2863–2872.
- Schutt, R., 2012. *Investigating the Social World : The Process and Practice of Research*, 7th ed. Sage Publications, Thousand Oaks, California.

- Sears, D.O., 1986. College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *J. Pers. Soc. Psychol.* 51, 515.
- Sheehan, K.B., 2001. E-mail Survey Response Rates: A Review. *J. Comput.-Mediat. Commun.* 6.
- Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S., 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Anal. Mach. Intell. IEEE Trans.* 31, 39–58.