A primer for conducting survey research using MTurk: Tips for the field

Silvana Chambers

Kim Nimon
University of Texas at Tyler, knimon@uttyler.edu

Paula Anthony-McMann

Follow this and additional works at: https://scholarworks.uttyler.edu/hrd_fac

Recommended Citation
http://hdl.handle.net/10950/2326

This Article is brought to you for free and open access by the Human Resource Development at Scholar Works at UT Tyler. It has been accepted for inclusion in Human Resource Development Faculty Publications and Presentations by an authorized administrator of Scholar Works at UT Tyler. For more information, please contact tgullings@uttyler.edu.
A Primer for Conducting Survey Research using MTurk: Tips for the Field

Silvana Chambers, University of Texas at Tyler, Tyler, TX, USA
Kim Nimon, University of Texas at Tyler, Tyler, TX, USA
Paula Anthony-McMann, ETMC Regional Health Care System, Tyler, TX, USA

ABSTRACT

This paper presents best practices for conducting survey research using Amazon Mechanical Turk (MTurk). Readers will learn the benefits, limitations, and trade-offs of using MTurk as compared to other recruitment services, including SurveyMonkey and Qualtrics. A synthesis of survey design guidelines along with a sample survey are presented to help researchers collect the best quality data. Techniques, including SPSS and R syntax, are provided that demonstrate how users can clean resulting data and identify valid responses for which workers could be paid.

KEYWORDS

MTurk, Qualtrics, Survey, SurveyMonkey

INTRODUCTION

Over the past 25 years, the Internet has progressively become part of how we live our lives. It has changed the way in which we communicate and exchange knowledge. Consequently, the Internet has become a tool for conducting academic and organizational research (Callegaro, Baker, Bethlehem, Goritz, Krosnick, & Lavrakas, 2014; Granello & Wheaton, 2004; Oppenheimer, Pannucci, Kasten, & Haase, 2011). In the field of human resource development, for example, is not uncommon for findings from survey research to inform theory and/or practice (cf. Gubbins & Rousse, 2015; Shuck & Reio, 2011).

Crowdsourcing is a technical innovation that refers to the process of obtaining content by soliciting contributions from a large pool of people, particularly from online communities. Recent studies have found that data collected through crowdsourcing were as good or better than data collected by more traditional survey methods (Behrend, Sharek, Meade, & Wiebe, 2011; Feitosa, Joseph, & Newman, 2015). One of the most popular crowdsourcing services used by social science researchers to recruit participants is MTurk (Buhrmester, Kwang, & Gosling, 2011).

In this paper, we offer a primer to researchers interested in using MTurk for data collection. First, we present MTurk and compare it to other survey participant recruitment services (i.e., SurveyMonkey,
Qualtrics). Second, we discuss the limitations and benefits associated with MTurk. Third, we synthesize best practices for designing a survey to be deployed on MTurk and present an associated example. Fourth, we review the implications of collecting data from MTurk and provide R and SPSS syntax that may be helpful starting places for researchers who are new to collecting data from MTurk.

**MTurk**

MTurk, short for Amazon Mechanical Turk, is a service offered by Amazon.com, Inc. It was originally designed for internal purposes in which participants would generate descriptors (e.g., “modern,” “vase,” “ivory”) for Amazon products, and in exchange, they received very small financial incentives (Landers & Behrend, 2015). Today, MTurk has evolved into a service that connects researchers (requestors) with respondents (workers) via Amazon’s online marketplace. MTurk recruits individuals for their marketplace by offering small researcher-paid financial incentives for HIT (Human Intelligence Tasks) completion (Buhrmester et al., 2011). MTurk gives researchers absolute discretion over the financial incentive their survey participants receive. Research on MTurk workers indicated that they are younger (millennials >50%), more educated (college degree >50%), and have lower income ($20,000–29,999) (cf. Berinsky, Huber, & Lenz, 2012; Feitosa et al., 2015; Smith, Roster, Golden & Albam, 2015) as compared to the average U.S. population. Therefore, MTurk workers may not be suitable for all studies.

Horton and Chilton (2010) reported that the hourly wage of the typical MTurk worker is $1.38. However, several studies (e.g., Casler, Bickel, & Hackett, 2013; Shapiro, Chandler, & Mueller, 2013) reported rates two to three times as high. MTurk adds a 40% commission to HIT’s total cost plus an additional 5% for Master workers, who are workers who have earned that qualification for consistently receiving positive feedback. Therefore, the cost per survey response for a 10-minute survey can range from $0.30 to $1.00.

**Other Recruitment Services**

SurveyMonkey and Qualtrics are two of the leading non-traditional survey recruitment and distribution services used in organizational, marketing, and academic research. SurveyMonkey and Qualtrics provide survey responses from their online panels in exchange for researcher-paid incentives/fees. SurveyMonkey and Qualtrics also provide online survey design software that allow individuals with basic computer skills to develop professional looking surveys in very little time (Brandon, Long, Loraas, Mueller-Phillips, & Vansant, 2013).

**SurveyMonkey**

SurveyMonkey Audience (SMA) is a service that connects researchers with SurveyMonkey’s online pool of members. Each SMA member completes a profile that includes demographic questions when joining a panel, which is used to identify potential participants based on the requested target population. Participants who complete surveys receive a $0.50–$1.00 donation to their charity of choice and, if offered by the researcher, a chance to participate in a sweepstake. The cost for the researcher varies based on the length of the survey. A 30-item survey costs around $3.00–$4.00 per response plus an optional, but recommended, sweepstake incentive of up to $500 (SurveyMonkey, n. d.). Studies that used SMA data have shown a wide range of usable response rates, including reported rates of 16 per cent (Morris, 2013) and 95 per cent (Cook & Lorass, 2013).
**Qualtrics**

Qualtrics offers a more personalized service to data collection in which an assigned Qualtrics project manager assists the researcher in exploring the options that best fit the researcher’s needs, which include but are not limited to identifying the best target demographics, survey length, and survey design. The cost to the researcher varies based on the length of the survey and target demographics. The base cost per response is $5.00 with a project minimum of $500 for a survey without specific targeting requirements. Specific targeting requirements (e.g., industry, employment type, educational attainment) increase the cost per response (Qualtrics, n. d.). Published studies have reported a wide range of usable response rates for data collected through Qualtrics, including rates of 20 per cent (Hazen, Kung, Cegielkski, & Jones-Farmer, 2014) and 90 per cent (Mohr, Lichtenstein, & Janiszewski, 2012).

**MTurk to Online Panels Comparison**

Consider a hypothetical example of collecting data from 300 U.S.-based, full-time working adults using a 40-item multiple choice survey that takes an average of seven minutes to complete. Given the hypothetical example, Table 1 depicts MTurk, SurveyMonkey, and Qualtrics’ services and costs. Like SurveyMonkey and Qualtrics online panels, MTurk allows researchers to screen participants based on their desired population target and provides a wide pool of potential participants. In comparison to SurveyMonkey and Qualtrics, MTurk is the most cost-effective option for conducting survey research.

<table>
<thead>
<tr>
<th>Table 1. Comparison of online participant recruitment services</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service</strong></td>
</tr>
<tr>
<td>Cost - Hypothetical Case</td>
</tr>
<tr>
<td>Cost per participant</td>
</tr>
<tr>
<td>Participant Incentive</td>
</tr>
<tr>
<td>Optional Incentives</td>
</tr>
<tr>
<td>Project Total Cost</td>
</tr>
<tr>
<td>Payment Terms</td>
</tr>
<tr>
<td>Software/Survey Design</td>
</tr>
<tr>
<td>Customer Support</td>
</tr>
</tbody>
</table>

Note: *U.S. based full-time employees (n=300). *Does not include optional incentives. *Cost obtained using SurveyMonkey’s online pricing tool. *Purchase of additional responses may be necessary to reach desired sample size. *Qualtrics does not publish online panel costs; costs were obtained through a call to their sales department. Information about services was retrieved from the SurveyMonkey, Qualtrics and MTurk websites.
In addition to a U.S. pool of participants, MTurk has a large international pool as well. However, despite the benefits of having a more diverse population, research (e.g., Berinsky et al., 2012; Paolacci, Chandler, & Ipeirotis, 2010) shows that having a large number of international participants makes MTurk samples less demographically representative of the U.S. population when compared with SurveyMonkey and Qualtrics. Additionally, participants from countries whose primary language is not English may have an effect on data quality (Feitosa et al., 2015). Nonetheless, MTurk allows researchers to screen participants’ location and nationality. Therefore, the risk of having international respondents can be mitigated in the survey design and deployment process.

Accounting for the differences previously noted, data quality and results obtained from MTurk samples are similar to SurveyMonkey and Qualtrics samples (Berinsky et al., 2012) as well as more traditional samples (Berinsky et al., 2012; Buhrmester et al., 2011; Horton, Rand, & Zeckhauser, 2011; Mason & Watts, 2009; Paolacci et al., 2010). Additionally, unlike SurveyMonkey and Qualtrics, MTurk allows researchers to approve or reject payment on the basis of data quality (Brandon et al., 2013).

**MTurk LIMITATIONS**

A major objection to the use of MTurk for academic research is the threat to external validity (Crowston, 2012). External validity is “the degree to which the results obtained in a given study would hold true at other times, in other settings, or with other individuals” (Sacket & Larson, 1990, p. 430). However, the external validity of data collected via MTurk appears to be an advantage rather than a concern as MTurk uniquely offers a wide range of diversity in the U.S. and internationally (Behrend et al., 2011; Buhrmester et al., 2011; Landers & Behrend, 2015). In fact, the notion that online panels threaten the validity of a study’s results is not based on empirical evidence, but is rather the result of an arbitrary categorization of data sources (Landers & Behrend, 2015). MTurk samples might be considered a threat to external validity only if the sample is not sufficiently similar to the intended population to draw conclusions that are generalizable to it (Yin, 2013).

Repeated participation is a concern that could threaten the internal validity of a study. However, it may not be a concern for organizational and human resource development (HRD) surveys in which participants are asked to share their opinions, feelings, or perceptions (cf. Chandler, Mueller, & Paolacci, 2014). While it is a concern when participants of a study are asked to perform cognitive tasks, it is perhaps reasonable to assume that participants may not change their opinions, feelings, or perceptions with more exposure to instrumentation (Landers & Behrend, 2015). Nevertheless, repeated participation of respondents in the same study should not be a concern as researchers using MTurk can design their HIT to prevent repeated participation (Chandler et al., 2014).

The use of financial incentives is a potential concern, as it could introduce systematic bias to a study (cf. Cobanoglu & Cobanoglu, 2003). However, research (Landers & Behrend, 2015) suggests that the payment MTurk workers receive for completing surveys does not seem to be a source of bias. Despite having an incentive for participation, MTurk samples seem to be no different than traditional samples (Hsieh & Kocielnik, 2016; Komarov, Reinecke, & Gajos, 2013) and yield similar levels of data quality (Behrend et al., 2011; Buhrmester et al., 2011; Chandler et al., 2014; Mason & Watts, 2009). The use of financial incentives as a source of bias should only be a concern if financial incentives are theoretically linked to outcomes (Jenkins, Gupta, Mitra, & Shaw, 1998; Landers & Behrend, 2015). Furthermore, research (Chandler et al., 2014; Rao, Kaminska, & McCutcheon, 2010) indicates that financial incentives increase participation rates without decreasing data quality.

High rates of non-response create a higher probability of statistical biases, which could represent a threat to the external validity of a study (Tomaskovic-Devey, Leiter, & Thompson, 1994). Consequently, response rates are of concern to scholars. A high response rate is associated with larger sample sizes, which result in higher statistical power and greater credibility (Rogelberg and Stanton, 2007). Therefore, to increase the precision of estimates and statistical power of a study, it
is recommended to increase sample size (Thompson, 2006). Unlike SurveyMonkey and Qualtrics, MTurk uses a marketplace in which researchers publish their surveys to be accessed by Amazon’s pool of workers. The unknown number of potential respondents who see the survey but choose not to participate makes it impossible to accurately calculate response rates. However, scholars argue that in survey research, response representativeness is more important than response rate. Therefore, response rate should be a concern only if it affects the representativeness of the sample (Cook, Heath, & Thompson, 2000).

Despite being unable to calculate response rates on initial studies, panel studies are available for subsequent studies. MTurk panel studies are done through the MTurk interface, which sends out an invite to all participants of a researcher’s previous study, thus, allowing a researcher to calculate response rates using the number of total surveys sent out. Research has found that response rates on MTurk for studies conducted eight days, three weeks, and three months after the initial survey were in the range of 60% to 68% (Berinsky et al., 2012; Buhrmester et al., 2011). For researchers using MTurk for initial studies, it is perhaps of more value to reiterate that large sample sizes can be achieved quickly. Berinsky et al. (2012) conducted several studies using MTurk for data collection and reported that at incentive levels of $0.75 and $0.50 per participant, they collected 400 completed surveys ($n = 400$) within one and three days, respectively (Berinsky et al., 2012).

Another source of concern is that MTurk gives researchers the option to approve or reject payment (Fleischer, Mead, & Huang, 2015; Paolacci et al., 2010). Although this can be seen as an advantage, it can also be considered a risk. The subjective nature of responses used for organizational and HRD studies can make it difficult to justify payment rejections for straight lining or for using neutral responses for all items. Furthermore, rejecting payment could potentially violate institutional review board (IRB) guidelines, as investigators must honor incentives offered to participants who complete, are disqualified, or are unable to complete a survey through no fault of their own (45 CFR 46). There are, however, techniques in survey HIT design that can reduce the risk of getting poor quality responses. MTurk records workers’ performance (HIT approval rate) and researchers have the option to design their HITs limiting participation of workers based on that criterion (e.g., 95% HIT approval rate).

**MTurk BENEFITS**

Despite skepticism about this relatively new tool for data collection, research provides evidence that MTurk: (a) offers an inexpensive and quick alternative to more traditional survey research methods, with advantages over other online panels (Brandon et al., 2013), (b) yields quality data that show measurement invariance with other methods (Berinsky et al., 2012; Casler et al., 2013; Feitosa et al., 2015; Paolacci et al., 2010), and (c) is appropriate for social sciences research (Behrend et al., 2011; Buhrmester et al., 2011; Landers & Behrend, 2015). As previously noted, concerns about MTurk can be mitigated with thoughtful research, survey, and HIT design.

An analysis of recent HRD academic literature revealed that, while still in the initial stages, MTurk samples are being used in published studies, more so than Qualtrics or SurveyMonkey samples (see Table 2). Like other traditional and non-traditional sampling methods, MTurk has advantages and disadvantages, which need to be evaluated against each specific research study. There are instances in which MTurk has advantages over traditional samples. MTurk workers are anonymous to researchers. Therefore, responses cannot be linked to identity, which represents a potential advantage for studies that involve sensitive questions. Institutional review boards are likely to treat studies that use MTurk participants as exempt from reviews (Paolacci et al., 2010). Finally, the particular demographics of the MTurk population (educated millennial) may represent an opportunity for HRD scholars. Millennials, who have redefined what being “social” means, are also known as “digital natives” (Prensky, 2001). Millennials continue to enter the workforce in large numbers and have been the subject of much
research due to their connection to technology that makes them communicate, learn, and exchange knowledge differently than older generations (Hershatter & Epstein, 2010).

As mentioned previously, MTurk’s capability of reaching international participants is an advantage for researchers who wish to collect data from participants outside of the U.S. In the same way, scholars are able to filter the location of workers to U.S. workers only under the “worker requirements” preference; they could broaden their participant pool to include participants worldwide or select a specific country. Chua (2013) explored the effects of ambient cultural disharmony on creative thinking using participants from five different countries. Mason and Watts (2009) examined the relationship between financial incentives and performance using a sample of participants from 43 different countries. Eriksson and Simpson (2010) explored the relationship between gender differences and risk aversion using samples from the U.S. and India. These are just a few examples of multiple studies that have taken advantage of MTurk’s international population.

**Table 2. Online data sources used in HRD articles based on a keyword search (2000–Present)**

<table>
<thead>
<tr>
<th>Journal</th>
<th>Journal of Management Journal</th>
<th>1,580</th>
<th>11</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Academy of Management Review</td>
<td>962</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Administrative Science Quarterly</td>
<td>631</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Adult Education Quarterly</td>
<td>307</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Adult Learning</td>
<td>330</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>British Journal of Management</td>
<td>733</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>European Journal of Training and Development, previously published as Journal of European Industrial Training</td>
<td>204</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Human Resource Development Quarterly</td>
<td>516</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Human Resource Development International</td>
<td>641</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>International Journal of Adult Vocational Education and Technology</td>
<td>76</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Journal of Applied Behavioral Science</td>
<td>369</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Journal of Business and Psychology</td>
<td>667</td>
<td>15</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Journal of European Industrial Training</td>
<td>506</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Journal of Human Resources</td>
<td>1,890</td>
<td>0</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Journal of Organizational Behavior</td>
<td>1,500</td>
<td>9</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>T+D (Training and Development)</td>
<td>288</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Work and Occupations</td>
<td>320</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Despite all the benefits associated with collecting data using MTurk (e.g., cost, convenience, large pool of potential participants, response rate), the use of proper techniques for survey design and data collection can impact response rate and data quality. While MTurk provides a survey development tool, there are other online applications that are more user friendly and provide more customization tools and functionality (Holden, Dennie, & Hicks, 2013). Some of the leading providers of online survey software are SurveyMonkey, Qualtrics, and Zoomerang. Qualtrics has been used in conjunction with
MTurk in several studies (Casler et al., 2013; Holden et al., 2013; Smith et al., 2015). Qualtrics allows users to create surveys and generate reports through a user-friendly GUI (graphical user interface) and distribute them through several channels, such as MTurk. Qualtrics can be used on PC and Mac computers, and works with most browsers. Additionally, Qualtrics provides an easy way to provide unique identifiers for each completed survey, which can be used for data verification for accepting and rejecting payment for participants.

An example survey developed for MTurk data collection using Qualtrics incorporating the recommendations offered in this article is available at https://uttyler.az1.qualtrics.com/SE/?SID=SV_afz2pTKyqYsRLHD. For the purpose of illustration, items in the surveyed are numbered. Note that the survey indicates official sponsorship by including the university name and logo on each page, which mitigates non-response bias (cf. Fan & Yan, 2010). The survey also includes a screening question (Q1 in example survey) and an informed consent page that may exit participants from the survey if screening criteria are not met or consent is not provided. For participants that meet the criteria of the study and complete the survey, a unique code is presented that participants can use when requesting payment.

Data safety and usability of the survey are factors that can have an effect on response rates (Fan & Yan, 2010). This risk is mitigated by: (a) using survey software (i.e., Qualtrics) that is accessible by different browsers and computer platforms (Couper, Tourangeau, Conrad & Crawford, 2004), (b) protecting data by using a unique (not shared) ID and password to log in and storing data on secure servers (Kraut, Olson, Banaji, Bruckman, Cohen, & Couper, 2004), (c) not collecting participant identifying data and explicitly informing participants of it (Rogelberg, Spitzmüller, Little, & Reeve, 2006) (d) formatting the survey in a clean, easy to read manner, using paging instead of scrolling (Peytchev, Couper, McCabe, & Crawford, 2006), (e) using simple, unambiguous language for instructions (Dillman & Smyth, 2007), and (f) ensuring that the survey does not take too long to complete, preferably less than 13 minutes (Handwerk, Carson & Blackwell, 2000).

Common method variance (CMV) has been identified as a potential problem associated with research in the social sciences, particularly that involving self-reporting instruments that use a single instrument for independent and dependent variables data collection (Reio, 2010; Richardson, Simmering, & Sturman, 2009). To minimize the risk of CMV, procedures for strengthening its procedural design should be implemented (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Reio, 2010). As seen in the example survey, directions for the items (Q4, Q5, and Q6–Q10) indicate that there are no right or wrong answers and that responses remain anonymous to avoid evaluation apprehension. As implemented in the example survey, it is helpful to divide the survey into blocks (e.g., dependent variable block, independent variable block, and demographics block). The dependent variable block (Q4 in example survey) should be placed before the independent variable block (Q5 in example survey) to avoid item priming effects (Podsakoff, et al., 2003). Scales within each block should be randomized or counterbalanced; however, it is not recommended that items from different scales are randomized indiscriminately, as it may disrupt the respondent's thought process and decrease the quality of data (Chang, Witteloostuijn, & Eden, 2010; Craighead, Ketchen, Dunn, & Hult, 2011; Podsakoff et al., 2003).

Despite the popularity of progress bars in online surveys, it is recommended that a progress bar not be used for MTurk data collection, as research (Villar, Callegaro, & Yang, 2013) shows that including a progress indicator in surveys that provide financial incentives is associated with increased drop-off rates. Conversely, faster completion rates have been associated with higher pay rates (i.e., $.75 for a 7–9 min. survey) (Berinksy et al., 2014). However, higher pay does not seem to have an effect on data reliability (Buhrmester et al., 2011; Mason & Watts, 2009; Rouse, 2015).

What could impact data quality when using MTurk is the potential risk of surveys being answered by artificial intelligence or “bots.” To mitigate the risks of unreliable data due to bot-generated survey responses, a screening question should be presented to ensure that the participant is human and has an understanding of the English language (e.g., “What is the third word in this question: How many stars
are in the American flag?"; Rouse [2015, p. 305]; Q2 in example survey). Additionally, to ensure that participants are paying attention to the survey, instructional manipulation checks (IMC) or attention checks could be included throughout the survey (e.g., “Please click on the little blue circle at the bottom of the screen. Do not click on the scale items that are labeled from 1 to 9”; Oppenheimer, Meyvis, Davidenko [2009, p. 871]; see item between Q3 and Q4 in example survey). These attention checks, as can be observed on the example survey provided, serve as “speed bumps” to help participants focus, or re-focus, on the survey, as well as to help the researcher identify responses that may be adding noise to the sample and decreasing the quality of the data (Oppenheimer et al., 2009; Smith et al., 2015). The researcher should consider the trade-off of eliminating responses of participants who fail the attention checks. While eliminating those responses can improve the quality of the data and increase statistical power, it could also harm the external validity of the study (Oppenheimer et al., 2009).

Finally, the decision of demographic questions (Q6 through Q10 in example survey) placement should be made considering the nature of the survey as well as the distribution channel. Organizational surveys have a higher perceived psychological risk due to a closer connection between participants and survey sponsors, which increases participants’ fear of being identified and fear of repercussion, thus increasing nonresponse and drop-out rates when demographic questions are placed at the beginning of a survey (Borg, Braun, & Baumgartner, 2008; Rogelberg, Conway, Sederburg, Spitzmuller, Aziz, & Knight, 2003). Therefore, it is recommended that demographic questions be placed at the end of the survey when they are organizational or sensitive in nature (e.g., medical history, sexual orientation) (Colton & Covert, 2007; Teclaw, Price, & Osatuke, 2012). However, when using MTurk, it is reasonable to assume that most participants do not have a close connection with the researcher that would create a fear of repercussions for their survey responses. Therefore, it is recommended that demographic questions in organizational or HRD research surveys distributed using MTurk not be treated with extra sensitivity due to their organizational nature since there is likely no connection between the researcher and participants.

MTurk DATA COLLECTION

Upon completion of designing an electronic survey, survey platforms (e.g., Qualtrics, SurveyMonkey) provide researchers with survey links that can be shared with prospective participants. The link is shared with the MTurk participant pool through the creating of a HIT. MTurk allows researchers to screen participants to fit their desired demographic criteria. Although research (Berinsky et al. 2012; Buhrmester et al., 2011; Paolacci et al., 2010) shows that MTurk samples can be as representative of the U.S. population as traditional participant pools (e.g., students, convenience), researchers must select the option of filtering the location of workers to “U.S. workers only” under the worker requirements preferences, in order to achieve quality results that are representative of the population (Feitosa et al., 2015), unless the researcher intends to include international participants in their study.

Although getting participants in MTurk is not a big challenge, it can be time consuming for researchers that offer small financial incentives for surveys that take more time to complete as compared to others in Amazon’s online marketplace. To mitigate that risk and improve response rates, a researcher might consider providing a higher pay rate (Berinsky et al, 2012; Rao et al., 2010). Note that higher pay rates have only been associated with faster response rates and do not appear to have an effect on data quality (Buhrmester et al., 2011; Mason & Watts, 2009; Rouse, 2015). Researchers also need to recognize that potential participants are informed of the incentive amount and estimated completion time by the researcher (see Figure 1). Therefore, if a survey is wrongfully advertised to have an estimated completion time of an hour and has an incentive of $0.50, response rates may be lower than if the researcher had specified a more realistic estimated completion time of 15 minutes.

As mentioned previously, one of the advantages of MTurk over other online panels is that researchers have the option to automatically approve, or to review and approve or reject payment to participants based on completion or response quality, which is recorded on each worker’s profile.
This can be done by using the unique code identifier generated by Qualtrics (or other survey software platform) to match the responses from MTurk (see last screen in example survey when meeting the study criteria and completing the survey). However, rejecting payment to workers should be the last route for a researcher, as it could violate IRB guidelines. Researchers could increase their chances of getting quality data designing HITs that limit participation of workers who have a HIT approval rate of 95 per cent, as it would indicate that they have a good reputation with other researchers (Barger & Sinar, 2011; Berinsky et al., 2012; Brandon et al., 2013). Another practice that can be used to improve data quality obtained through MTurk is awarding bonuses, which can only be done by using unique survey codes. The use of bonuses for good performance is a way to get more diligent responses, which not only encourages workers to provide good responses to get a bonus payment, but tacitly lets them know that the data will be reviewed prior to releasing payment (Barger & Sinar, 2011).

Finally, it is imperative to let participants know that the data will be inspected for quality and only complete surveys will receive a unique code that can claim payment. This is done through the HIT instructions provided to the MTurk community (see Figure 1). Instructions should include language that explicitly informs workers that only those who pass the screening questions will be able to take the survey, which upon completion will generate a unique code that workers can use to get credit for the survey. Additionally, the researcher can inform workers about bonus offers and that the data will be checked for quality.

An example of syntax for data cleaning is included in Appendix A (R) and Appendix B (SPSS), which can be used to match the data file exported from Qualtrics to the data file from MTurk. This is done through the computation of a variable that identifies potentially unusable responses for deletion, such as the length of time of survey completion, straight lining, failed attention checks, and failed bot checks. The use of that feature simplifies the process of approving/rejecting payment; however, researchers should use their best judgment when deciding what to eliminate from the sample.
CONCLUSION

While we have presented a compilation of best practices for survey research using MTurk, we only addressed limitations associated with survey design and deployment. MTurk’s strengths and weaknesses need to be evaluated as there is no one-fits-all solution when it comes to sampling approaches. However, despite its limitations, research (Buhrmester et al., 2011; Berinsky et al., 2012; Landers & Berend, 2015) appears to support the use of MTurk as a viable alternative for scholars. One of MTurk’s major advantages, as compared to other non-traditional participant recruitment methods, is that it may help overcome one of the barriers to data collection—access to quality data at low cost.
REFERENCES


PMID:14992637


Silvana Chambers is a PhD student of human resource development at The University of Texas at Tyler, Department of Human Resource Development, 3900 University Blvd, Tyler, TX 75799. Her current research interests are statistical methodologies and diversity in organizations.

Kim Nimon, PhD, is an associate professor of human resource development at The University of Texas at Tyler, Department of Human Resource Development, 3900 University Blvd, Tyler, TX 75799. Her research currently focuses on employee engagement and quantitative research methodologies.

Paula Anthony-McMann, PhD, is the Vice President and CIO of ETMC Regional Healthcare System, 1000 South Beckham Avenue, Tyler, TX 75701. She is also the President and CEO of HealthFirst, an ETMC subsidiary.
APPENDIX A

Sample R Syntax to Clean Data and Prepare for MTurk Payment
### Change to your working directory
setwd("C:/Users/MTurk/Survey")

### Load libraries
library(foreign, pos=4)
library(psych)

### Read in dataset (one version with coded values and the other as choice text
dso1 <-
  read.table("SurveyValues.csv", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
dso2 <-
  read.table("SurveyText.csv", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)

### Look at dataset and column ids
head(dso2)
names(dso2)

### Create dataset with coded values
ds<-dso1

### Overwrite demographics and screening questions with data from choice text file
ds[,c(11,12,51:55)]<-dso2[,c(11,12,51:55)]

### Change names of demographic columns
names(ds)[51:55]<-c("Gender","Cohort","Mgr","Dept","Other")

### See total responses
nrow(ds)

### Initialize delete variable
ds$Delete<"Keep"

### Flag responses that did not pass screening questions
table(ds$Employed, ds$WorkHours)
levels(ds$Employed)
levels(ds$WorkHours)

ds$Delete[(ds$Employed=="No")|(ds$WorkHours!="40 or more")]<"Screen"

### Flag responses from BOTs
table(ds$BotCheck, useNA="ifany")
ds$Delete<"Keep" & (ds$BotCheck!=4)<"BOT"

### Flag responses that did not pass screening questions

### Flag responses that did not consent

table(ds$Consent, useNA = "ifany")

```r
ds$Delete[(ds$Delete == "Keep") & is.na(ds$Consent)] <- "Consent"
table(ds$Delete)
```

### Flag responses that did not pass IMC1

table(ds$IMC1_1, useNA = "ifany")

```r
ds$Delete[(ds$Delete == "Keep") & !is.na(ds$IMC1_1)] <- "IMC1"
table(ds$Delete)
```

### Flag responses that did not pass IMC2

table(ds$IMC2, useNA = "ifany")

```r
ds$Delete[(ds$Delete == "Keep") & is.na(ds$IMC2)] <- "IMC2"
table(ds$Delete)
```

### Flag incompleters

ds$Delete[(ds$Delete == "Keep") & (ds$V10 == 0)] <- "Incomplete"
table(ds$Delete)

### Create variable that shows elapsed time of survey

d$sTime <- as.numeric(strptime(ds$V9, "%m/%d/%Y %H:%M") - strptime(ds$V8, "%m/%d/%Y %H:%M"))/60

d$sTime <- as.numeric(strptime(ds$V9, "%m/%d/%Y %H:%M") - strptime(ds$V8, "%m/%d/%Y %H:%M"))

### Look at dataset to see if Time is created appropriately
### May need to divide Time by 60 seconds to get minutes or not. Select appropriate code above.

```r
#edit(ds)
```

### Flag duration <5 minutes > 60 minutes

ds$Delete[(ds$Delete == "Keep") & ((ds$sTime < 5) | (ds$sTime > 60))] <- "Time"
table(ds$Delete)

### Create variable that shows standard deviation of how people respond to PANAS items

ds$PANASsd <- apply(subset(ds, select = interested:afriad), 1, sd)

### Create variable that shows standard deviation of how people respond to UWES items

ds$UWESsd <- apply(subset(ds, select = VI1:AB5), 1, sd)

### Flag straight lined responses to DVs, IVs, duration <5 minutes > 60 minutes, or incompleters

ds$Delete[(ds$Delete == "Keep") & ((ds$PANASsd == 0) | (ds$UWESsd == 0))]

### Write dataset out that can be used to assist determining MTurk payment

write.csv(ds, "SurveyOrig.csv", row.names = FALSE)

### Omit unusable responses

ds <- subset(ds, Delete == "Keep")
nrow(ds)
describe(subset(ds,select=c(interested:afriad,VI1:AB5)))

table(ds$Gender)
table(ds$Gender)/nrow(ds)

table(ds$Cohort)
table(ds$Cohort)/nrow(ds)

table(ds$Mgr)
table(ds$Mgr)/nrow(ds)

table(ds$Dept)
table(ds$Dept)/nrow(ds)

table(ds$Other)

write.csv(ds,"SurveyClean.csv",row.names=FALSE)
APPENDIX B

Sample SPSS Syntax to Clean Data and Prepare for MTurk Payment
CD ‘C:\Users\Survey\Desktop\Survey’.

USE ALL.
COMPUTE Delete=0.
EXECUTE.

VARIABLE LABELS
Delete.

VALUE LABELS
Delete
0 ‘Keep’
1 ‘Screen’
2 ‘Bot’
3 ‘Consent’
4 ‘IMC1’
5 ‘IMC2’
6 ‘INC’
7 ‘Time’
8 ‘Straight’.
EXECUTE.

FREQUENCIES VARIABLES=V1
/FORMAT=NOTABLE
/ORDER=ANALYSIS.

CROSSTABS
/TABLES=Employed BY WorkHours
/FORMAT=AVALUE TABLES
/CELLS=COUNT
/COUNT ROUND CELL.

IF ((Employed ~= 1) | (WorkHours ~= 5)) Delete=1.

FREQUENCIES VARIABLES=Delete
/ORDER=ANALYSIS.

FREQUENCIES VARIABLES=BotCheck
/ORDER=ANALYSIS.

IF ((Delete = 0) & (BotCheck ~=4)) Delete = 2.

FREQUENCIES VARIABLES=Delete
/ORDER=ANALYSIS.

FREQUENCIES VARIABLES=Consent
/ORDER=ANALYSIS.

IF ((Delete = 0) & (Consent ~=1)) Delete = 3.

FREQUENCIES VARIABLES=Delete
/ORDER=ANALYSIS.

FREQUENCIES VARIABLES=IMC1_1
/ORDER=ANALYSIS.

IF ((Delete = 0) & (Missing(IMC1_1) = 0)) Delete = 4.

FREQUENCIES VARIABLES=Delete
/ORDER=ANALYSIS.

FREQUENCIES VARIABLES=IMC2
/ORDER=ANALYSIS.

IF ((Delete = 0) & (Missing(IMC2) = 1)) Delete = 5.

FREQUENCIES VARIABLES=Delete
/ORDER=ANALYSIS.

FREQUENCIES VARIABLES=V10
/ORDER=ANALYSIS.

IF ((Delete = 0) & (V10 = 0)) Delete = 6.

FREQUENCIES VARIABLES=Delete
/ORDER=ANALYSIS.

COMPUTE Time = RND(V9/60)-RND(V8/60).
EXECUTE.

DESCRIPTIVES VARIABLES=Time
/STATISTICS=MEAN STDDEV MIN MAX.

IF ((Delete = 0) & ((Time < 5) | (Time > 60))) Delete=7.
EXECUTE.

FREQUENCIES VARIABLES=Delete
/ORDER=ANALYSIS.

COMPUTE PANASSD=SD(interested, distressed, excited, upset, strong, guilty, scared, hostile, enthusiastic, proud, irritable, alert, ashamed, inspired, nervous, determined, attentive, jittery, active, afraid).
EXECUTE.

COMPUTE UWESSD=SD(VI1,VI2,DE2,DE3,VI3,AB3,DE4,AB4,AB5).
EXECUTE.
IF MISSING(PANASSD) PANASSD=0.
EXECUTE.

IF MISSING(UWESSD) UWESSD=0.
EXECUTE.

IF ((Delete =0) & ((PANASSD = 0) | (UWESSD = 0))) Delete=8.
EXECUTE.

FREQUENCIES VARIABLES=Delete
/ORDER=ANALYSIS.

SAVE OUTFILE='SurveyOrigUpdated.sav'
/COMPRESSED.

FILTER OFF.
USE ALL.
SELECT IF (Delete=0).
EXECUTE.

FREQUENCIES VARIABLES=V1
/FORMAT=NOTABLE
/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=interested distressed excited upset strong guilty scared hostile enthusiastic proud irritable alert ashamed inspired nervous determined attentive jittery active afraid V11 V12 DE2 DE3 V13 AB3 DE4 AB4 AB5
/STATISTICS=MEAN STDDEV MIN MAX.

FREQUENCIES VARIABLES=Gender GenCohort Supervisor Department Department_TEXT
/ORDER=ANALYSIS.

SAVE OUTFILE='SurveyClean.sav'
/COMPRESSED.