A beginner's guide and best practices for using crowdsourcing platforms for survey research: The case of Amazon Mechanical Turk (MTurk)

Cihan Cobanoglu
University of South Florida, CIHAN@CIHAN.ORG

Muhittin Cavusoglu
Northern Arizona University, muhittin@mcavusoglu.com

Gozde Turktarhan
University of South Florida, gozdeturktarhan@gmail.com

A beginner's guide and best practices for using crowdsourcing platforms for survey research: The case of Amazon Mechanical Turk (MTurk)

Authors

Corresponding Author

Cihan Cobanoglu, University of South Florida Sarasota–Manatee, School of Hospitality and Tourism Management, 8350 N Tamiami Trail, Sarasota, FL 34243

Abstract

Introduction

Researchers around the globe are utilizing crowdsourcing tools to reach respondents for quantitative and qualitative research (Chambers & Nimon, 2019). Many social science and business journals are receiving studies that utilize crowdsourcing tools such as Amazon Mechanical Turk (MTurk), Qualtrics, MicroWorkers, ShortTask, ClickWorker, and Crowdsource (e.g., Ahn, & Back, 2019; Ali et al., 2021; Esfahani, & Ozturk, 2019; Jeong, & Lee, 2017; Zhang et al., 2017). Even though the use of these tools presents a great opportunity for sharing large quantities of data quickly, some challenges must also be addressed. The purpose of this guide is to present the basic ideas behind the use of crowdsourcing for survey research and provide a primer for best practices that will increase their validity and reliability.

What is crowdsourcing research?

Crowdsourcing describes the collection of information, opinions, or other types of input from a large number of people, typically via the internet, and which may or may not receive (financial) compensation (Hargrave, 2019; Oxford Dictionary, n.d.). Within the behavioral science realm, crowdsourcing is defined as the use of internet services for hosting research activities and for creating opportunities for a large population of participants. Applications of crowdsourcing techniques have evolved over the decades, establishing the strong informational power of crowds. The advent of Web 2.0 has expanded the possibilities of crowdsourcing, with new online tools such as online reviews, forums, Wikipedia, Qualtrics, or MTurk, but also other platforms such as Crowdflower and Prolific Academic (Peer et al., 2017; Sheehan, 2018).

Crowdsourcing platforms in the age of Web 2.0 use remote labor recruited via the internet to assist employers complete tasks that cannot be left to machines. Key characteristics of crowdsourcing include payment for workers, their recruitment from any location, and the completion of tasks (Behrend et al., 2011). They also allow for a relatively quick collection of data compared to data collection in the field, and participants are rewarded with an incentive—often financial compensation. Crowdsourcing not only offers a large participation pool but also a streamlined process for the study design, participant recruitment, and data collection as well as integrated participant compensation system (Buhrmester et al., 2011). Also, compared to other traditional marketing firms, crowdsourcing makes it easier to detect possible sampling biases (Garrow et al., 2020). Due to advantages such as reduced costs, diversity of participants, and flexibility, crowdsourcing platforms have surged in popularity for researchers.

Advantages

MTurk is one of the most popular crowdsourcing platforms among researchers, allowing *Requesters* to submit tasks for *Workers* to complete (Cummings & Sibona, 2017). MTurk has been used as an online crowdsourcing platform for the recruitment of human subjects for research purposes (Paolacci & Chandler, 2014). Research has also shown MTurk to be a reliable and cost-effective tool, capable of

providing representative data for research in the behavioral sciences (e.g., Crump et al., 2013; Goodman et al., 2013; Mason & Suri, 2012; Rand, 2012; Simcox & Fiez, 2014). In addition to its use in social science studies, the platform has been used in marketing, hospitality and tourism, psychology, political science, communication, and sociology contexts (Sheehan, 2018). To illustrate, between 2012 and 2017, more than 40% of the studies published in the *Journal of Consumer Research* used crowdsourcing websites for their data collection (Goodman & Paolacci, 2017).

Disadvantages

Although researchers have assessed crowdsourcing platforms as reliable and cost-effective for data collection in the behavioral sciences, they are not exempt of flaws. One disadvantage is the possibility of unsatisfactory data quality. In fact, the virtual setting of the survey implies that the investigator is physically separated from the participant, and this lack of monitoring could lead to data quality issues (Sheehan, 2018). In addition, participants in survey research on crowdsourcing platforms are not always who they claim to be, creating issues of trust with the data provided and, ultimately, the quality of the research findings (McGonagle, 2015; Smith et al., 2016).

A recurrent concern with MTurk workers, for instance, is their assessment as experienced survey takers (Chandler et al., 2015). This experience is mainly acquired through completion of dozens of surveys per day, especially when they are faced with similar items and scales. Smith et al. (2016) identified two types of problems performing data collection using MTurk; namely, cheaters and speeders. As compared to Qualtrics—which has a strict screening and quality-control processes to ensure that participants are who they claim to be—MTurk appears to be less exigent regarding the workers. However, a downside for data collection with Qualtrics is more expensive fees—about $5.00 per questionnaire on Qualtrics, against $0.50 to $1.50 on MTurk (Ford, 2017). Hence, few researchers were able to conduct surveys and compare respondent pools with Qualtrics or other traditional marketing research firms (Garrow et al., 2020).

Another challenge using MTurk arises when trying to collect a desired number of responses from a population targeted to a specific city or area (Ross et al., 2010). The issues inherent to the selection process of MTurk have been the subject of investigations in several studies (e.g., Berinsky et al., 2012; Chandler et al., 2014; 2015; Harms & DeSimone, 2015; Paolacci et al., 2010; Rand, 2012). Feitosa et al. (2015) pointed out that international respondents may still identify themselves as U.S. respondents with the use of fake addresses and accounts. They found that 5% to 10% of participants identifying themselves as U.S. respondents were actually from overseas locations. Moreover, Babin et al. (2016) assessed that the use of trap questions allowed researchers to uncover that many respondents change their genders, ages, careers, or income within the course of a single survey. The issues of (a) experienced workers for the quality control of questions and (b) speeders, which, for MTurk can be attributed to the platform being the main source of revenue for a given respondent, remain the inherent issues of crowdsourcing platforms used for research purposes.

Best practices

Some best practices can be recommended in the use of crowdsourcing platforms for data collection purposes. Workers IDs can be matched with IDs from previous studies, thus allowing researchers to exclude responses from workers who had answered previous similar studies (Goodman & Paolacci, 2017). Furthermore, proceed to a manual assignment of qualification on MTurk prior to data collection (Litman et al., 2015; Park & Park, 2020). When dealing with experienced workers, both using multiple attention checks and optimizing the survey in a way to have the participants exposed to the stimuli for a sufficient length of time to better address the questions are also recommended (Sheehan, 2018). In this sense, shorter surveys are preferred to longer ones, which affect the participant's concentration, and may, in turn, adversely impact the quality of their answers. Most importantly, pretest the survey to make sure

that all parts are working as expected.

Researchers should also keep in mind that in the context of MTurk, the primary method for measurement is the web interface. Thus, to avoid method biases, researchers should ponder whether or not method factors emerge in the latent measurement models (Podsakoff et al., 2012). As such, time-lagged research designs may be preferred as predictor and criterion variables can be measured at different points in time or administered in different platforms, such as Qualtrics vs MTurk (Cheung et al., 2017). In general, the use of crowdsourcing platforms including MTurk may be appropriate according to the research question; and the quality of data is reliant on the quality-control strategies used by researchers to enhance data quality. Trade-offs between various validity types need to be prioritized according to the research objectives (Cheung et al., 2017).

From our experience using crowdsourcing tools for our own research as the editorial team members of several journals and chair of several conferences, we provide the best practices as outlined below:

MTurk Worker (Respondent) Selection:

- Researchers should consider their study population before using MTurk for data collection. The MTurk platform should be used for the appropriate study population. For example, if the study targets restaurant owners or company CEOs, MTurk workers may not be suitable for the study. However, if the target population is diners, hotel guests, grocery shoppers, online shoppers, students, or hourly employees, utilizing a sample from MTurk would be suitable.
- Researchers should use the selection tool in the software. For example, if you target workers only from one country, exclude responses that came from an internet protocol (IP) address outside the targeted country and report the results in the method section.
- Researchers should consider the demographics of workers on MTurk which must reflect the study targeted population. For example, if the study focuses on baby boomers use of technology, then the MTurk sample should include only baby boomers. Similarly, the gender balance, racial composition, and income of people on MTurk should mirror the targeted population.
- Researchers should use multiple screening tools that identify quality respondents and avoid problematic response patterns. For example, MTurk provides the approval rate for the respondents. This refers to how many times a respondent is rejected for various reasons (i.e., wrong code entered). We recommend using a 90% or higher approval rate.
- Researchers should include screening questions in different places with different type of questions to make sure that the respondents are appropriate for your study. One way is to use knowledge-based questions about the subject. For example, rather than asking "How experienced are you with accounting practices?", a supplemental question such as "Which of the following is a component of an income statement?" should be integrated into the study in a different section of the survey.

Survey Validity:

- Researchers should conduct a pilot survey from MTurk workers to identify and fix any potential data quality and programming problems before the entire data set is collected. Researcher can estimate time required to complete the survey from the pilot study. This average time should be used in calculating incentive payment for the workers in such a way that the payment should equate or exceed minimum wage in the targeted country.
- Researchers should build multiple validity-check tools into the survey. One of them is to ask attention check questions such as "please click on 'strongly agree' in this question" or "What is 2+2? Please choose 5" (Cobanoglu et al., 2016) Even though these attention questions are good and should be implemented, experienced survey takers or bots easily identify them and answer them correctly, but then give random answers to other questions. Instead, we recommend building in more involved validity check questions. One of the best is asking the same question in different places and in different forms. For example, asking the age of the respondent in the beginning of the survey and then asking them the year of their birth at the end of the survey is an effective way to check that they are replying to the survey honestly. Exclude all those who answered the same question differently. Report the results of these validity checks in the methodology. Cavusoglu (2019) found that almost 20% of the surveys were eliminated due to the failure of the validity check questions which were embedded in different places and in different forms in his survey.
- Researchers should be aware of *internet bot,* which is a software that runs automated tasks. Some respondents use a bot to reply to the surveys. To avoid this, use Captcha verification, which forces respondents to perform random tasks such as moving the bar to a certain area, clicking in boxes that has cars, or checking boxes to verify the person taking the survey is not a bot.
- Whenever appropriate, researchers should use time limit options offered by online survey tools such as Qualtrics to control the time that a survey taker must spend to advance to the next question. We found that this is a great tool, especially when you want the respondents to watch a video, read a scenario, or look at a picture before they respond to other questions.
- Researchers should collect data in different days and times during the week to collect a more diverse and representative sample.

Data Cleaning:

- Researchers should be aware that some respondents do not read questions. They simply select random answers or type nonsense text. To exclude them from the study, manually inspect the data. Exclude anyone who filled out the survey too quickly. We recommend excluding all responses filled out less than 40% of the average time to take the survey. For example, if it takes 10 minutes to fill out a survey, we exclude everyone who fills out this

survey in 4 minutes or less. After we separated these two groups, we compared them and found that the speeders' (aka cheaters) data was significantly different than the regular group.

- Researchers should always collect more data than needed. Our rule of thumb is to collect 30% more data than needed. For example, if 500 clean data responses are wanted, collect at least 650 data. The targeted number of data will still be available after cleaning the data. Report the process of cleaning data in the method section of your article, showing the editor and reviewers that you have taken steps to increase the validity and reliability of the survey responses.

- Calculating a response rate for the samples using MTurk is not possible. However, it is possible to calculate active response rate (Ali et al., 2021). It can be calculated as the raw response numbers deducted from all screening and validity check question results. For example, if you have 1000 raw responses and you eliminated 100 responses for coming from IP address outside of the United States, another 100 surveys for failing the validity check questions, then your active response rate would be 800/1000= 80%.

Keywords
data collection, validity-check questions, incentive, survey, online data collection, survey screening questions

Revisions
Cobanoglu, C., Cavusoglu, M., & Turktarhan, G. (2021). *A beginner's guide and best practices for using crowdsourcing platforms for survey research: The case of Amazon Mechanical Turk (MTurk)* [Editorial]. Journal of Global Business Insights, 6(1), 92-97. https://www.doi.org/10.5038/2640-6489.6.1.1177

Creative Commons License

A Beginner's Guide and Best Practices for Using Crowdsourcing Platforms for Survey Research: The Case of Amazon Mechanical Turk (MTurk)

Cihan Cobanoglu[1], Muhittin Cavusoglu[2], and Gozde Turktarhan[3]

Muma College of Business
[1]cihan@cihan.org
[3]gozdeturktarhan@gmail.com

The W. A. Franke College of Business
Northern Arizona University
[2]muhittin@mcavusoglu.com

Introduction

Researchers around the globe are utilizing crowdsourcing tools to reach respondents for quantitative and qualitative research (Chambers & Nimon, 2019). Many social science and business journals are receiving studies that utilize crowdsourcing tools such as Amazon Mechanical Turk (MTurk), Qualtrics, MicroWorkers, ShortTask, ClickWorker, and Crowdsource (e.g., Ahn, & Back, 2019; Ali et al., 2021; Esfahani, & Ozturk, 2019; Jeong, & Lee, 2017; Zhang et al., 2017). Even though the use of these tools presents a great opportunity for sharing large quantities of data quickly, some challenges must also be addressed. The purpose of this guide is to present the basic ideas behind the use of crowdsourcing for survey research and provide a primer for best practices that will increase their validity and reliability.

What is crowdsourcing research?

Crowdsourcing describes the collection of information, opinions, or other types of input from a large number of people, typically via the internet, and which may or may not receive (financial) compensation (Hargrave, 2019; Oxford Dictionary, n.d.). Within the behavioral science realm, crowdsourcing is defined as the use of internet services for hosting research activities and for creating opportunities for a large population of participants. Applications of crowdsourcing techniques have evolved over the decades, establishing the strong informational power of crowds. The advent of Web 2.0 has expanded the possibilities of crowdsourcing, with new online tools such as online reviews, forums, Wikipedia, Qualtrics, or MTurk, but also other platforms such as Crowdflower and Prolific Academic (Peer et al., 2017; Sheehan, 2018).

Crowdsourcing platforms in the age of Web 2.0 use remote labor recruited via the internet to assist employers complete tasks that cannot be left to machines. Key characteristics of crowdsourcing include payment for workers, their recruitment from any location, and the completion of tasks (Behrend et al., 2011). They also allow for a relatively quick collection of data compared to data collection in the field, and participants are rewarded with an incentive—often financial compensation. Crowdsourcing not only offers a large participation pool but also a streamlined process for the study design, participant recruitment, and data collection as well as integrated participant compensation system (Buhrmester et al., 2011). Also, compared to other traditional

marketing firms, crowdsourcing makes it easier to detect possible sampling biases (Garrow et al., 2020). Due to advantages such as reduced costs, diversity of participants, and flexibility, crowdsourcing platforms have surged in popularity for researchers.

Advantages

MTurk is one of the most popular crowdsourcing platforms among researchers, allowing *Requesters* to submit tasks for *Workers* to complete (Cummings & Sibona, 2017). MTurk has been used as an online crowdsourcing platform for the recruitment of human subjects for research purposes (Paolacci & Chandler, 2014). Research has also shown MTurk to be a reliable and cost-effective tool, capable of providing representative data for research in the behavioral sciences (e.g., Crump et al., 2013; Goodman et al., 2013; Mason & Suri, 2012; Rand, 2012; Simcox & Fiez, 2014). In addition to its use in social science studies, the platform has been used in marketing, hospitality and tourism, psychology, political science, communication, and sociology contexts (Sheehan, 2018). To illustrate, between 2012 and 2017, more than 40% of the studies published in the *Journal of Consumer Research* used crowdsourcing websites for their data collection (Goodman & Paolacci, 2017).

Disadvantages

Although researchers have assessed crowdsourcing platforms as reliable and cost-effective for data collection in the behavioral sciences, they are not exempt of flaws. One disadvantage is the possibility of unsatisfactory data quality. In fact, the virtual setting of the survey implies that the investigator is physically separated from the participant, and this lack of monitoring could lead to data quality issues (Sheehan, 2018). In addition, participants in survey research on crowdsourcing platforms are not always who they claim to be, creating issues of trust with the data provided and, ultimately, the quality of the research findings (McGonagle, 2015; Smith et al., 2016).

A recurrent concern with MTurk workers, for instance, is their assessment as experienced survey takers (Chandler et al., 2015). This experience is mainly acquired through completion of dozens of surveys per day, especially when they are faced with similar items and scales. Smith et al. (2016) identified two types of problems performing data collection using MTurk; namely, cheaters and speeders. As compared to Qualtrics—which has a strict screening and quality-control processes to ensure that participants are who they claim to be—MTurk appears to be less exigent regarding the workers. However, a downside for data collection with Qualtrics is more expensive fees—about $5.00 per questionnaire on Qualtrics, against $0.50 to $1.50 on MTurk (Ford, 2017). Hence, few researchers were able to conduct surveys and compare respondent pools with Qualtrics or other traditional marketing research firms (Garrow et al., 2020).

Another challenge using MTurk arises when trying to collect a desired number of responses from a population targeted to a specific city or area (Ross et al., 2010). The issues inherent to the selection process of MTurk have been the subject of investigations in several studies (e.g., Berinsky et al., 2012; Chandler et al., 2014; 2015; Harms & DeSimone, 2015; Paolacci et al., 2010; Rand, 2012). Feitosa et al. (2015) pointed out that international respondents may still identify themselves as U.S. respondents with the use of fake addresses and accounts. They found that 5% to 10% of participants identifying themselves as U.S. respondents were actually from overseas locations. Moreover, Babin et al. (2016) assessed that the use of trap questions allowed researchers

to uncover that many respondents change their genders, ages, careers, or income within the course of a single survey. The issues of (a) experienced workers for the quality control of questions and (b) speeders, which, for MTurk can be attributed to the platform being the main source of revenue for a given respondent, remain the inherent issues of crowdsourcing platforms used for research purposes.

Best practices

Some best practices can be recommended in the use of crowdsourcing platforms for data collection purposes. Workers IDs can be matched with IDs from previous studies, thus allowing researchers to exclude responses from workers who had answered previous similar studies (Goodman & Paolacci, 2017). Furthermore, proceed to a manual assignment of qualification on MTurk prior to data collection (Litman et al., 2015; Park & Park, 2020). When dealing with experienced workers, both using multiple attention checks and optimizing the survey in a way to have the participants exposed to the stimuli for a sufficient length of time to better address the questions are also recommended (Sheehan, 2018). In this sense, shorter surveys are preferred to longer ones, which affect the participant's concentration, and may, in turn, adversely impact the quality of their answers. Most importantly, pretest the survey to make sure that all parts are working as expected.

Researchers should also keep in mind that in the context of MTurk, the primary method for measurement is the web interface. Thus, to avoid method biases, researchers should ponder whether or not method factors emerge in the latent measurement models (Podsakoff et al., 2012). As such, time-lagged research designs may be preferred as predictor and criterion variables can be measured at different points in time or administered in different platforms, such as Qualtrics vs MTurk (Cheung et al., 2017). In general, the use of crowdsourcing platforms including MTurk may be appropriate according to the research question; and the quality of data is reliant on the quality-control strategies used by researchers to enhance data quality. Trade-offs between various validity types need to be prioritized according to the research objectives (Cheung et al., 2017).

From our experience using crowdsourcing tools for our own research as the editorial team members of several journals and chair of several conferences, we provide the best practices as outlined below:

MTurk Worker (Respondent) Selection:

- Researchers should consider their study population before using MTurk for data collection. The MTurk platform should be used for the appropriate study population. For example, if the study targets restaurant owners or company CEOs, MTurk workers may not be suitable for the study. However, if the target population is diners, hotel guests, grocery shoppers, online shoppers, students, or hourly employees, utilizing a sample from MTurk would be suitable.
- Researchers should use the selection tool in the software. For example, if you target workers only from one country, exclude responses that came from an internet protocol (IP) address outside the targeted country and report the results in the method section.
- Researchers should consider the demographics of workers on MTurk which must reflect the study targeted population. For example, if the study focuses on baby boomers use of technology, then the MTurk sample should include only baby boomers. Similarly, the

gender balance, racial composition, and income of people on MTurk should mirror the targeted population.

- Researchers should use multiple screening tools that identify quality respondents and avoid problematic response patterns. For example, MTurk provides the approval rate for the respondents. This refers to how many times a respondent is rejected for various reasons (i.e., wrong code entered). We recommend using a 90% or higher approval rate.
- Researchers should include screening questions in different places with different type of questions to make sure that the respondents are appropriate for your study. One way is to use knowledge-based questions about the subject. For example, rather than asking "How experienced are you with accounting practices?", a supplemental question such as "Which of the following is a component of an income statement?" should be integrated into the study in a different section of the survey.

Survey Validity:

- Researchers should conduct a pilot survey from MTurk workers to identify and fix any potential data quality and programming problems before the entire data set is collected. Researcher can estimate time required to complete the survey from the pilot study. This average time should be used in calculating incentive payment for the workers in such a way that the payment should equate or exceed minimum wage in the targeted country.
- Researchers should build multiple validity-check tools into the survey. One of them is to ask attention check questions such as "please click on 'strongly agree' in this question" or "What is 2+2? Please choose 5" (Cobanoglu et al., 2016) Even though these attention questions are good and should be implemented, experienced survey takers or bots easily identify them and answer them correctly, but then give random answers to other questions. Instead, we recommend building in more involved validity check questions. One of the best is asking the same question in different places and in different forms. For example, asking the age of the respondent in the beginning of the survey and then asking them the year of their birth at the end of the survey is an effective way to check that they are replying to the survey honestly. Exclude all those who answered the same question differently. Report the results of these validity checks in the methodology. Cavusoglu (2019) found that almost 20% of the surveys were eliminated due to the failure of the validity check questions which were embedded in different places and in different forms in his survey.
- Researchers should be aware of *internet bot,* which is a software that runs automated tasks. Some respondents use a bot to reply to the surveys. To avoid this, use Captcha verification, which forces respondents to perform random tasks such as moving the bar to a certain area, clicking in boxes that has cars, or checking boxes to verify the person taking the survey is not a bot.
- Whenever appropriate, researchers should use time limit options offered by online survey tools such as Qualtrics to control the time that a survey taker must spend to advance to the next question. We found that this is a great tool, especially when you want the respondents to watch a video, read a scenario, or look at a picture before they respond to other questions.
- Researchers should collect data in different days and times during the week to collect a more diverse and representative sample.

Data Cleaning:

- Researchers should be aware that some respondents do not read questions. They simply select random answers or type nonsense text. To exclude them from the study, manually inspect the data. Exclude anyone who filled out the survey too quickly. We recommend excluding all responses filled out less than 40% of the average time to take the survey. For example, if it takes 10 minutes to fill out a survey, we exclude everyone who fills out this survey in 4 minutes or less. After we separated these two groups, we compared them and found that the speeders' (aka cheaters) data was significantly different than the regular group.

- Researchers should always collect more data than needed. Our rule of thumb is to collect 30% more data than needed. For example, if 500 clean data responses are wanted, collect at least 650 data. The targeted number of data will still be available after cleaning the data. Report the process of cleaning data in the method section of your article, showing the editor and reviewers that you have taken steps to increase the validity and reliability of the survey responses.

- Calculating a response rate for the samples using MTurk is not possible. However, it is possible to calculate active response rate (Ali et al., 2021). It can be calculated as the raw response numbers deducted from all screening and validity check question results. For example, if you have 1000 raw responses and you eliminated 100 responses for coming from IP address outside of the United States, another 100 surveys for failing the validity check questions, then your active response rate would be 800/1000= 80%.

References

Ahn, J., & Back, K. J. (2019). Cruise brand experience: Functional and wellness value creation in tourism business. *International Journal of Contemporary Hospitality Management, 31*(5), 2205-2223.

Ali, F., Ciftci, O., Nanu, L., Cobanoglu, C., & Ryu, K. (2021). Response rates in hospitality research: An overview of current practice and suggestions for future research. *Cornell Hospitality Quarterly, 62*(1), 105-120.

Ali, F., Terrah, A., Wu, C., Ali, L., & Wu, H. (2021). Antecedents and consequences of user engagement in smartphone travel apps. *Journal of Hospitality and Tourism Technology*. Advance online publication. https://doi.org/10.1108/JHTT-09-2020-0221

Babin, B. J., Griffin, M., & Hair Jr, J. F. (2016). Heresies and sacred cows in scholarly marketing publications. *Journal of Business Research, 69*(8), 3133-3138.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*(3), 800-813.

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.

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.

Cavusoglu, M. (2019). *Online and self-directed learning readiness among hospitality and tourism college students and industry professionals* (Publication No. 13805377) [Doctoral dissertation, University of South Florida]. ProQuest Dissertations & Theses Global.

Cobanoglu, C., Berezina, K., Cavusoglu, M., & Ali, F. (2016, May 9-13). *Using Mechanical Turk respondents as a replacement of general public in online surveys* [Conference Presentation]. International Interdisciplinary Business-Economics Advancement Conference, Miami, FL.

Chambers, S., & Nimon, K. (2019). Conducting survey research using MTurk. In M. Khosrow-Pour, S. Clarke, M. E. Jennex, A. Becker, & A. V. Anttiroiko (Eds.), *Crowdsourcing: Concepts, methodologies, tools, and applications* (Vol. 3., pp. 410-439). IGI Global.

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.

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using nonnaive participants can reduce effect sizes. *Psychological Science, 26*(7), 1131–1139.

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology, 32*(4), 347-361.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One, 8*(3), Article e57410. https://doi.org/10.1371/journal.pone.0057410

Cummings, J., & Sibona, C. (2017). Crowdsourcing surveys: Alternative approaches to survey collection. *Journal of Information Systems Applied Research, 10*(1), 44-54.

Esfahani, S. S., & Ozturk, A. B. (2019). The influence of individual differences on NFC-based mobile payment adoption in the restaurant industry. *Journal of Hospitality and Tourism Technology, 10*(2), 219-232.

Feitosa, J., Joseph, D., & Newman, D. (2015). Crowdsourcing and personality measurement equivalence: A warning about countries whose primary language is not English. *Personality and Individual Differences*, *75*(4), 47-52.

Ford, J. B. (2017). Amazon's Mechanical Turk: A comment. *Journal of Advertising, 46*(1), 156-158.

Garrow, L. A., Chen, Z., Ilbeigi, M., & Lurkin, V. (2020). A new twist on the gig economy: Conducting surveys on Amazon Mechanical Turk. *Transportation*, *47*(1), 23-42.

Goodman, J. K., & Paolacci, G. (2017). Crowdsourcing consumer research. *Journal of Consumer Research, 44*(1), 196-210.

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.

Hargrave, M. (2019). *Crowdsourcing*. Investopedia. https://www.investopedia.com/terms/c/crowdsourcing.asp

Harms, P., & DeSimone, J. (2015). Caution! MTurk workers ahead – Fines doubled. *Industrial and Organizational Psychology*, *8*(2), 183-190.

Jeong, M., & Lee, S. A. (2017). Do customers care about types of hotel service recovery efforts? *Journal of Hospitality and Tourism Technology, 8*(1), 5-18.

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US-and India-based workers on Mechanical Turk. *Behavior Research Methods, 47*(2), 519-528.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*(1), 1-23.

McGonagle, A. K. (2015). Participant motivation: A critical consideration. *Industrial and Organizational Psychology, 8*(2), 208-214.

Oxford Dictionary. (n.d.). Crowdsourcing. In *Oxford learner's dictionaries*. Retrieved February 25, 2021, from https://www.oxfordlearnersdictionaries.com/us/definition/english/crowdsourcing

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184-188.

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

Park, H., & Park, E. (2020). How to enhance tourist perceptions of environmental issues through nature images: An importance-performance analysis. *Journal of Global Business Insights, 5*(2), 120-133.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*(3), 153-163.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.

Rand, D. G. (2012). The promise of Amazon Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, *299*(8), 172–179.

Ross, J., Irani, I., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. In E. Mynatt, & D. Schoner (Eds.), *CHI 2010 extended abstracts on human factors in computing systems* (pp. 2863-2872). Association for Computing Machinery.

Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs, 85*(1), 140-156.

Simcox, T., & Fiez, J. A. (2014). Collecting response times using Amazon Mechanical Turk and Adobe Flash. *Behavior Research Methods, 46*(1), 95-111.

Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research, 69*(8), 3139-3148.

Zhang, T. C., Omran, B. A., & Cobanoglu, C. (2017). Generation Y's positive and negative eWOM: Use of social media and mobile technology. *International Journal of Contemporary Hospitality Management, 29*(2), 732-761.