

MECHANICAL TURK (MTURK) FOR ONLINE RESEARCH

This guidance document is intended for investigators conducting online research using Amazon's Mechanical Turk (MTurk), based on the most up-to-date information regarding MTurk as of January 2020. Should you need additional assistance, please contact OPHS at 510-642-7461 or ophs@berkeley.edu.

Table of Contents:

- A. Introduction**
- B. What Is MTurk?**
- C. Is MTurk Anonymous?**
 - 1. Internal HITs**
 - 2. External HITs**
 - 3. NotifyWorkers**
- D. MTurk and General Data Protection Regulation (GDPR)**
- E. MTurk and Informed Consent**
- F. Glossary of Terms**
- G. References**

A. Introduction

Use of Amazon's Mechanical Turk (MTurk) has become increasingly popular for conducting online research involving surveys, as MTurk facilitates access to a large and diverse participant population at a relatively low cost to investigators. This guidance is intended to provide investigators with helpful suggestions for addressing issues of privacy and confidentiality that are specific to the MTurk marketplace, in order to conduct online research in accordance with [45 CFR 46](#).

For more information, please also review [CPHS Guidelines on Internet-Based Research](#).

B. What Is MTurk?

MTurk functions as a virtual labor market where registered Workers complete online Human Intelligence Tasks (HITs) for pay. During registration, all MTurk Workers are required to electronically sign a Participation Agreement confirming that they are at least 18 years of age. Likewise, individuals must register as MTurk Requesters to post HITs and collect data from consenting Workers.

MTurk [User Registration](#), [Terms of Service](#), and [Participation Agreement](#) should be read in their entirety.

C. Is MTurk Anonymous?

1. Internal HITs

For internal HITs, MTurk provides a template for the construction of surveys run directly on Amazon (Mason & Suri, 2011). For the following reasons, data may be identifiable

when HITs are internally run:

(a) *MTurk automatically collects IP addresses.* [MTurk's Privacy Policy](#) states that IP addresses are automatically collected and analyzed by Amazon. In general, CPHS considers IP addresses to be potentially identifiable, depending on context.

(b) *Amazon has access to Workers' personal identifiable information (PII) and survey responses.* Amazon has access to Workers' account information – such as name, email address, and physical address – and survey responses (Mason & Suri, 2011).

(c) *MTurk Worker IDs are automatically linked to survey responses.* [MTurk's Terms of Service \(ToS\)](#) do not allow for Requesters to directly access any PII from Workers. Consequently, all payment transactions are done using a 14-character alphanumeric code that is unique and specific to each MTurk Worker, and does not offer any clue as to a Worker's identity (Lease et al., 2013). This code, referred to as an MTurk Worker ID, is linked to survey responses for purposes of remuneration.

While Requesters do not have direct access to Workers' PII on MTurk, at least [one study](#) has shown that Worker IDs are linked to product reviews, ratings, and PII on publicly-available Amazon Profiles. This is because Amazon assigns one ID number to individuals who register for both online shopping and MTurk using the same account. For example:

Jane Doe creates an Amazon account for online shopping and uses the same email address to register for MTurk. Amazon assigns Jane the same alphanumeric ID number for both purposes: MT8936XJUB9352.

This ID number is embedded in the URL of users' Profile pages. Therefore, for MTurk survey respondents who also maintain a publicly-available Amazon Profile, it may be possible to identify individuals based on the amount and type of information shared on his/her/their public profile.

2. External HITs

Requesters may choose to use MTurk as a recruitment tool, but embed a link within MTurk that redirects Workers to complete online surveys via a separate, external site managed by the Requester, such as Qualtrics. For Workers who choose to complete external HITs, survey data are never available to Amazon MTurk. CPHS may consider this type of survey data collection to be generally anonymous (meaning that no identifiers can be linked to the data, either directly or through a coding system) given the following criteria are met:

- a. Survey questions do not ask for any PII or MTurk Worker IDs;
- b. It would not be possible to identify an individual based on his/her/their survey responses;
- c. IP addresses are not collected by the external site (e.g., Qualtrics); and
- d. Any compensation code automatically generated by the external site for payment in MTurk cannot be used to link individual MTurk Worker IDs to survey responses (i.e. each participant receives the same completion code at the end of the survey instead of a unique completion code linked to survey responses).

External HITs are recommended for MTurk survey research, particularly when questions are intended to solicit information that may be considered sensitive or personal, or that may pose greater-than-minimal risk to participants. However, investigators should keep in mind that Internet-based research can increase potential risks to confidentiality due to the possibility of third-party interception, and that no guarantees of complete confidentiality or anonymity can or should be made to study participants.

There are numerous publicly-available online tutorials for administering Qualtrics surveys via MTurk.

3. NotifyWorkers

As noted previously, MTurk does not allow for Requesters to directly access any PII from Workers. In the event that a Requester needs to contact an individual Worker or group of Workers, he/she may do so using MTurk's [NotifyWorkers](#) operation. This operation allows Requesters to send email to Workers using only Worker IDs. However, responding to a Requester's email will expose the Worker's email address to the Requester, and potentially his/her/their name along with information about the specific Task (e.g., survey).

Investigators should be wary of the possibility of exposing PII when using MTurk's NotifyWorkers operation.

D. MTurk and General Data Protection Regulation (GDPR)

[The General Data Protection Regulation](#) (GDPR) is a European Union (EU) regulation on data protection and privacy that went into effect on May 25, 2018. GDPR applies to all individuals **located** in the European Economic Area (EEA). Thus, whenever UC Berkeley investigators will collect or use Personal Data from research subjects located in the EEA, UC Berkeley is required to comply with GDPR requirements.

GDPR uses the term "Personal Data" to refer to any information relating to an individual who can be directly or indirectly identified. Examples of Personal Data include (but are not limited to):

- First and last name
- Email address
- ID number
- Location information such as home address or GPS coordinates
- Online identifiers such as IP address, browser or Internet cookie ID
- Phone number
- Demographic, behavioral or health-related information that could identify a person

Whenever data will be collected from MTurk subjects located in the EEA in a way that is not anonymous, GDPR requires that (a) specific information be provided to subjects regarding collection and use of their Personal Data, and (b) for [certain activities](#), explicit consent from subjects be obtained. UC Berkeley investigators are advised to include GDPR-mandated information in the informed consent form(s) to be presented to subjects. CPHS/OPHS has developed [template language](#) for investigators to use.

For more information, please also review [CPHS/OPHS guidance to General Data Protection Regulation \(GDPR\)](#).

E. MTurk and Informed Consent

Informed consent for Exempt-level MTurk research should follow [CPHS Guidelines for Exempt Research](#) for Respect for Persons and Informed Consent. Whenever Personal Data will be collected from subjects located in the EEA, investigators must also confirm in their Exempt application that they will provide subjects with CPHS/OPHS [template GDPR consent language](#) during the consent process.

For both Exempt and non-Exempt research, investigators should use the template consent forms for research available on the [CPHS website](#). In addition, the following information should be provided to MTurk subjects during the consent process:

If you agree to participate, please note that the data you provide may be collected and used by Amazon as per its privacy agreement. Additionally, if you are not currently located in the United States [if limiting the study to subjects in the U.S.] and/or are under the age of 18, please do not complete this [survey/HIT].”

Investigators may choose to use “attention checks,” also referred to as “trap questions,” meant to filter out data from subjects who are not attentive to the research questions. In many cases, failure to respond correctly to “attention check” questions results in the subject being excluded from further participation in the research. Furthermore, exclusion from further participation may result in the subject not receiving payment for completing a portion of the task.

For studies employing the use of “attention checks,” investigators should provide subjects with the following information during consent:

This study contains a number of checks to make sure that participants are finishing the tasks honestly and completely. As long as you read the instructions and complete the tasks, your HIT will be approved. If you fail these checks, your HIT will be rejected.

F. Glossary of Terms

Confidentiality: Pertains to the treatment of information that an individual has disclosed in a relationship of trust, and with the expectation that it will not be divulged to others without permission in ways that are inconsistent with the understanding of the original disclosure.

Human Intelligence Tasks (HITs): On MTurk, HITs are tasks posted by Requesters and completed by Workers, for pay.

Internet Protocol (IP) address: A numeric address assigned to every computer that connects to a network, or more commonly, the Internet.

Investigator: Any individual who contributes in a substantive way to the design, conduct, and/or analysis of the data of a study at or on behalf of the University of California, Berkeley.

Personal Data (GDPR): Any information relating to an identified or identifiable [natural person](#).

Privacy: Control over the extent, timing, and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others.

Requester: An individual who is registered on MTurk to post HITs for data collection (e.g., an Investigator or a Researcher).

Worker: An individual who is registered on MTurk to complete HITs for pay (e.g., a Participant or a Subject).

G. References

[General Data Protection Regulation \(GDPR\) – Official Legal Text](#), Retrieved January 21, 2020.

Lease, M., Hullman, J., Bigham, J.P., Bernstein, M., Kim, J., Lasecki, W.S., Bakhshi, S., Mitra, T., & Miller, R.C. (2013). [Mechanical Turk is not anonymous](#). *Social Science Research Network (online)*. Retrieved May 31, 2016.

Mason, W. & Suri, S. (2011). [Conducting behavioral research on Amazon’s Mechanical Turk](#). *Behavioral Research Methods*, 44(1), 1-23. Retrieved May 31, 2016.

Princeton, Survey Research Center, [“Using Amazon’s Mechanical Turk Service at Princeton.”](#) Retrieved May 31, 2016.