

Use of Amazon Mechanical Turk (MTurk) for Social/Behavioral Research Projects

Background:

Amazon Mechanical Turk (MTurk) is a crowd-sourcing Internet marketplace enabling individuals and businesses to coordinate the use of human intelligence to perform tasks (hence *Human Intelligence Tasks*, or HITs) that computers are currently unable to perform. MTurk has become a popular tool used in recruiting human subjects for social and behavioral research.

Things to Consider When Using MTurk

Q1: What should researchers know about Mechanical Turk?

First, researchers who plan to use MTurk should acquaint themselves with the service by visiting Amazon.com and reading about MTurk features.

- [Amazon's Advanced "How To" Guide:](#)
- [Amazon's FAQs:](#)

Q2: What privacy and confidentiality concerns does the use of MTurk raise?

Amazon.com has stated that the MTurk platform is NOT meant to support participant anonymity. MTurk Worker IDs are linked to the workers' Amazon.com public profiles.

Amazon.com may disclose worker information. Additionally, Worker information may be available to others (who submit a request) for tax reporting purposes.

Researchers should familiarize themselves with the privacy and confidentiality limitations associated with using MTurk. Amazon's [Privacy Policy](#).

"Information You Give Us" and *"Automatic Information,"* which outlines information internally associated with MTurk Worker data such as internet protocol (IP) address, social security number, bank account information, etc.

A 2013 study by researchers from the University of Texas at Austin and five other universities discovered how easy it is to uncover many Workers' personally identifying information. See: <https://news.utexas.edu/2013/03/27/amazons-online-workforce-not-so-anonymous-after-all>. Researchers should design their study to minimize privacy risks to participants.

Care should be taken as to the information that is requested and presented in the HITs. HITs can be reported by Workers for violating Amazon’s Acceptable Use policies or containing broken links or links that are not functioning properly. For details on Amazon’s policies and restrictions, see <https://www.mturk.com/acceptable-use-policy>.

Other examples of policy violations include:

- collecting personally identifiable information, including email addresses;
- generating referred site visits or “click-through” traffic;
- requiring Workers to register at another website or group;
- asking workers to take action to manipulate a website’s behavior or results;
- promoting a site, service or opinion;
- asking workers to solicit third parties;
- showing explicit or offensive content without the Adult Content Qualification; or
- requiring software that contains malware, spyware, viruses or other harmful code.

Q3: What information should be included in the IRB submission when the research involves the use of MTurk for recruiting subjects?

The study submission should describe the use of MTurk for subject recruitment, and include a statement in the “privacy and confidentiality” section that MTurk worker IDs will not be shared with anyone outside of the research team, will be removed from the data set, and/or will not be linked to survey/study responses (as applicable).

The HIT task description that will be posted in MTurk should be uploaded as part of the protocol submission, under the recruitment methods section of the myIRB SmartForm.

The protocol should also confirm that no Worker who has completed the task will be rejected and/or denied compensation. If there are any terminations prior to completion of the task, describe how that will be handled, for instance, by a lower “base level” of compensation, and/or instructions on closing the survey/browser window and "returning" the HIT. If there is the possibility of partial or no compensation, this should be stated in the informed consent.

Q4: What information should be included in the HIT description?

The HIT description to be posted to MTurk should include a brief description of the task involved, the compensation, any specific age range, and any specific inclusion criteria. However, since Amazon will know

which studies a worker has accessed for consideration, and might use that information for internal marketing and “user experience enhancement,” you should avoid any information in the HIT title or description indicating inclusion criteria that might be considered sensitive.

If your HIT could contain explicit or offensive content, for instance, nudity, the HIT should include the “Adult Content Qualification,” and the HIT Title following phrase: "(WARNING: This HIT may contain adult content. Worker discretion is advised.)"

Researchers should also take privacy concerns into consideration when developing IRB materials, especially the HIT description. For instance, the HIT description could include a disclaimer similar to this:

- Before you accept the HIT, please keep in mind that Amazon tracks the HITs that Workers click on, even if they ultimately decline to participate or withdraw after enrolling and don't get compensated. Therefore, do not accept this HIT if the study topic is of a sensitive nature and you wouldn't want your interest in the study disclosed to MTurk/Amazon.

Assuming the task proper is external to MTurk, researchers may include the statement:

- Your responses will not be accessible to Amazon as the survey is administered at a website external to Amazon MTurk.

Q5: What language should be included in the informed consent for studies involving the use of MTurk?

Suggestions to consider for Informed Consent Procedures and/or Study Documents:

- Add a disclaimer statement that any work performed on MTurk can be linked to the user's public profile page. Thus, Workers may wish to restrict what information they choose to share in their public profile.
- Add a statement in the consent document explaining that their MTurk Worker IDs (i.e., the 14-character sequence of letters and numbers used to identify Workers) will NOT be shared with anyone.
- If there will be study compensation, add a statement in the consent document that MTurk worker IDs will only be collected for the purposes of distributing compensation and will not be associated with survey responses (if applicable).
- There should be a section devoted to the length of time that the data will be stored or accessible by the researchers. Typically, the data will be stored for one to five years and destroyed after that.
- Since in almost all cases, the HIT itself is administered at an external website (e.g., Qualtrics), the consent should also include information regarding privacy and data security of that external site.

Q6: I want to screen participants for certain characteristics before letting them proceed to the main task of interest. How can this be done?

The simplest way is to have the desired characteristic(s) listed in the HIT title or description as inclusion criteria. This is equivalent to listing inclusion criteria in an ad or recruitment poster. Participants then self-exclude from the study if they don't qualify. (This is one of the most common and efficient approaches to inclusion/exclusion.)

Another approach is to state the criteria in the consent that participants will see after accepting the HIT and following the link to the study. The criteria can be bold-faced at the outset of the consent, with instructions to close the survey and/or browser window, and "return" or remove the HIT from their assignment list.

MTurk also provides a way to restrict HIT visibility on the basis of a wide range of what they call "qualifications," which can be anything from state of residence to education level to handedness, political affiliation, and education level (the two most commonly requested are gender and employment). Fees for this service can be significant.

The most common qualifications are *Approved HITs* and *Approval Rating*. An approval rating of >99% is essentially a 100% approval rating on MTurk. Only new workers who have completed a handful of tasks would qualify for this. The larger the number of approved HITs (>100, >1000, >5000, >10000) the more experienced the respondent, thus researchers can expect the quality of work to be higher.

Unpaid "screening surveys" are not uncommon on MTurk but should be limited to four or five questions. Simply end the survey if the criteria for the full survey are not met and ask the respondent to return the HIT. Returning the HIT does not influence their ability to participate in future HITs.

However, unpaid and unannounced screeners usually elicit a negative response from respondents and should be avoided; the possibility of unpaid termination should be described in the consent.

Under some circumstances, it might be appropriate to create your own "screening" survey as a preliminary or separate task, with appropriate compensation. (A good rule of thumb for pay is \$.10 cents a minute or \$6 an hour. This is the minimum amount that most experienced workers currently will accept.) Eligible respondents can be assigned a "qualification" that will allow them to take the full survey at a later date. The possibility of exclusion or further work, and compensation procedures, should be described in the HIT description or consent.

Q7: I need to "disguise" the inclusion criteria (e.g., some attitude or personal attribute), and include in the full task only those that qualify, but the basis for inclusion needs to be kept from workers, because it may bias their behavior in some important ways. How should this be done?

You can create a "basic" task as the first activity in the HIT. The critical screening question(s) can be embedded in additional (potentially irrelevant) material to disguise their importance for inclusion. All those that complete this "screening" HIT but who don't qualify for inclusion can be given compensation appropriate for the time taken. Those that qualify for the full task continue on (Qualtrics and other survey

hosts have mechanisms for such response-based conditional branching), and receive the basic compensation and a “bonus” for completing the full task.

The HIT description and consent could indicate that there is a short and long version of the HIT that they will be randomly assigned to (or, more transparently, that they will take the short version, which will determine if they qualify for the longer version, as appropriate) and indicate the compensation/bonus for the two versions.

Note that the same “basic compensation + bonus(es)” procedure can be used even if the screening criteria are not disguised, but may be described in the consent, to allow for different levels of compensation for different amounts of completed work within the same HIT.

From Amazon Web Services help:

- *If you want to offer Workers a reward within a range of amounts, you can post the HIT with a reward equal to the lowest amount in the range, then mention the offer of a bonus payment in the description of the HIT. For example, to offer a HIT with a reward between \$1 and \$5, post the HIT with a reward of \$1, then mention in the description that you will pay a bonus of up to \$4.*

Q8: Can I reject Workers after they have completed the study activity?

First, be sure to give Workers plenty of time to complete the task dutifully and honestly. Setting a six-minute timer on a six-minute survey forces Workers to rush. Most small surveys have a one-hour timer or more.

While researchers are free to discard data when “Workers” fail certain “attention checks” or respond so fast that the diligence given to the study task is questionable, UF IRB policy is that Workers who have consented to participate and completed the study activities from their perspective should NOT be “Rejected” as Workers, but compensated fully. (Note that this is more restrictive than Amazon’s own policy, which allows requesters to reject work deemed unsatisfactory.)

Rejection without compensation amounts to a “penalty for withdrawal” from the study, which is prohibited by the Common Rule. Moreover, a Rejection within MTurk lowers the Worker’s “approval rating,” which may prevent them from participating in future studies that have a minimum “rating” requirement. It is difficult to recover from a Rejection, so that is an additional penalty.

Finally, Rejections affect the Requester’s own statistics, and could make it more difficult for the researcher to acquire Workers in the future. More broadly, any such unilateral Rejections can “poison the waters” for other UF researchers, as Workers communicate extensively with each other on several forums.

Workers should not be rejected even if there are conditional terminations (because of a failed attention check response, for example) within Qualtrics or a similar survey platform. Once they've consented and started the HIT, they shouldn't be rejected for any reason. From the Worker’s perspective, they’ve “completed” the activity whenever Qualtrics terminates it. In such cases, the compensation + bonus

procedure described above could be used if there is a need for early termination in the task.

Final Thoughts about Rejections

An unfortunate side effect of the MTurk format, as with other online modes of interacting with participants, is that the “human factor” can easily get lost. Keep in mind that there are people on the other end providing your data, people from all walks of life with their own experiences.

The bottom line for this and other actions taken by researchers to exclude participants from any stated compensation or benefits is that participants may “withdraw consent” and stop participating at any time “with no penalty,” and a Rejection would be inconsistent with that statement. More broadly, though, the trust that is the basis of any human-subjects research is important to maintain. Two of the three basic ethical principles guiding such research are “respect and justice” for participants; without it, UF’s research enterprise couldn’t endure.

It’s worth remembering that the problem of participants not doing your activities in earnest is not unique to MTurk or other online platforms, but is made more salient by the commercial nature of MTurk. You should consider how you would handle the problem by analogy to other situations where you’re compensating participants for a task but don’t want to (fully) compensate those who “misbehave” or who ultimately fail to qualify for participation.

Summary of Sample Language

HIT Description

- This study involves *[insert description of study and task to be performed]*. To participate in this study, you must be *[include criteria such as age, location, other inclusion/exclusion criteria]*. You will be paid *[insert amount]* upon completion and approval of your work.
- Before you accept the HIT, please keep in mind that Amazon tracks the HITs that workers click on, even if they ultimately decline to participate or withdraw after enrolling and don't get compensated. Therefore, do not accept the HIT if the study topic is of a sensitive nature and you wouldn't want your interest in the study disclosed to MTurk/Amazon.

If the HIT contains explicit or offensive content include the following in the HIT title Per Amazon’s policy:

- (WARNING: This HIT may contain adult content. Worker discretion is advised.)

If the task is external to MTurk (i.e. a survey on Qualtrics), including the following:

- Your responses will not be accessible by Amazon as the *[task]* is administered at a website external to Amazon MTurk.

Informed Consent

In the confidentiality section of the informed consent include the following:

- Any work you perform on MTurk can be linked to your user's public profile page. You may wish to restrict what information you choose to share in your public profile.
- If you contact the Requester with comments or questions, your email address will automatically be inserted into the message so the Requester can reply to you.
- Worker IDs will be collected by the researchers for purposes of approving the HIT after completion, but will not be associated with your survey responses. MTurk Worker IDs will not be shared with anyone else.

If there is an explicit screening at the beginning of the task:

- There will be *[number of screening questions] [interspersed]* at the beginning of the *[task]*. We will check your eligibility based on your answer and only qualified Workers will proceed to the *[task]*. If you are not qualified, the *[task]* will terminate after the screening and you will *[not be compensated / receive [amount]]*. If you are qualified, you will be compensated *[amount]* after completion of the *[task]*.