

Questioning the Turk: Conducting High Quality Research with Amazon Mechanical Turk

#mturkacr

Joseph K. Goodman

Washington University in St. Louis

Gabriele Paolacci

Erasmus University Rotterdam

Game Plan and Ground Rules

- Have a conversation about Mturk
- Ask questions at any time
- Share your opinions
- Will post slides on “the blog”

[http://
experimentalturk.word
press.com/](http://experimentalturk.wordpress.com/)



The screenshot shows a blog post from 'Experimental Turk' titled 'MTurk workshop at ACR'. The post is dated October 14, 2014, and is written by Gabriele Paolacci. The main text discusses the increasing use of Mturk by behavioral researchers and the workshop's goal of questioning the Turk. It mentions a workshop at the North American ACR conference in Baltimore, MD, on October 25, 2014. The post includes social media sharing options (Twitter, Facebook, Email, Print), a 'Like' button, and a 'Leave a Comment' link. The right sidebar contains a 'SUBSCRIBE' section with RSS links, a 'CATEGORIES' list (conferences, games, heuristics and biases, information processing, papers, resources, Uncategorized), and a 'RECENT POSTS' list.

Experimental Turk
A blog on social science experiments on Amazon Mechanical Turk

About | How to collaborate | People | Resources | What is AMT

MTurk workshop at ACR | Posted by: Gabriele Paolacci | October 14, 2014

The use of Mturk by behavioral researchers continues to increase. Despite the evidence on the benefits (and drawbacks) of Mturk, many researchers, reviewers, and editors intuitively distrust the reliability and validity of online labor markets.

On October 25 , we will host a workshop at ACR called "Questioning the Turk: Conducting High Quality Research with Amazon Mechanical Turk". We will answer and debate questions from the ACR community regarding MTurk, and raise some new questions. We will discuss the current issues that arise from MTurk's use, as well as some of the solutions and replications. Questions can be submitted using the hashtag #mturkacr via Twitter (@aconsres, @joekgoodman, @gpaolacci) or Facebook (ACR page), as a comment to this post, or via email the organizers Joseph Goodman and Gabriele Paolacci.

The North American ACR conference will take place on October 24-26 at the Hilton Baltimore in Baltimore, MD. **The MTurk workshop will take place on Saturday, October 25, 2pm in the room Key 5.**

Share this:

★ Like

Be the first to like this.

Leave a Comment

Posted in conferences | Tags: mturk conferences

SUBSCRIBE

- RSS - Posts
- RSS - Comments

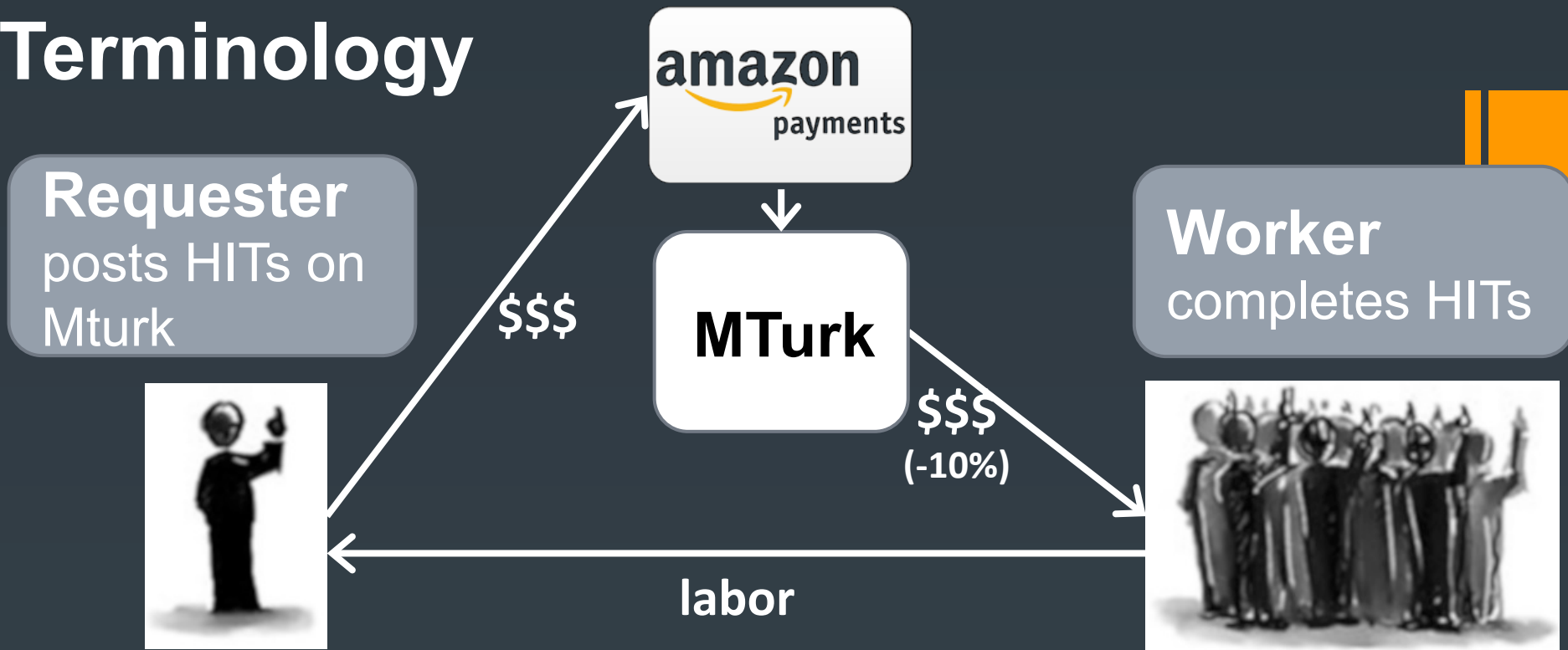
CATEGORIES

- conferences
- games
- heuristics and biases
- information processing
- papers
- resources
- Uncategorized

RECENT POSTS

- MTurk workshop at ACR
- Review of MTurk as a Participant Pool
- 2nd Workshop on Crowdsourcing and Online Behavioral Experiments
- Quick survey about software of online data collection
- Reputation as a Sufficient Condition for High Data Quality on MTurk
- MTurk Roundtable at ACR
- Consequences of Worker Nonnaivete: The Cognitive Reflection Test
- How naive are MTurk workers?
- TurkGate
- Using MTurk to Study Clinical Populations

Terminology



- **“HITs”** (Human Intelligence Tasks) tasks for money (\$\$.01 - \$64 a HIT)
- **Amazon Payments** – used to deposit \$ in and out of Mturk. Mturk holds requesters’ funds until payment is made to workers
- **Qualtrics** – online survey software (unrelated)

Who are these workers?

Diverse Group

- 500,000+ workers, 100+ countries
- As of today, 389,737 HITs available

- Previous research

(Goodman et al. 2013; Paolacci et al., 2010)

- Monday Survey

- w/ 95% approval filter
- \$.15 for 2 min survey



	<u>Previous Research</u>	<u>Monday's Survey</u>
Age	31	32
Male	58%	62%
<u>ESL</u>	<u>56%</u>	<u>6%</u>
<u>US</u>	<u>47%</u>	<u>93%</u>
<u>India</u>	<u>33%</u>	<u>6%</u>
Student	24%	21%
Some Univ	92%	94%
Bach Degree or higher		49%
Median Income		\$35,000
Employed Full-Time		48%
Employed Part-Time		20%
Ave # HITs per day		36.6
Mean # yrs on Mturk		1.37

Who are these workers?

Representative?

- Similar **income** distribution, slightly **younger**, fewer **children** compared to US population (Ipeirotis, 2010)

Spend a **day or less** per week on HITs (Ipeirotis, 2010)

Motivation?

- **Intrinsic** and **extrinsic**. Supplemental income, enjoyment (Paolacci et al. 2010)

What do they do on these HITs?

- Web searches, registering for websites, viewing ads, edit transcripts, test software, and fill out **SURVEYS!**



Is MTurk reliable?

- Answer is probably “**yes**”, but reliable for what?
- Right question is “**when**”...
- **Psychometrically Sound** (Buhrmester , Kwang, & Gosling, 2011)
- **Honest, consistent, and conscientious** (Rand, 2012; Shapiro et al., 2013)

Is MTurk reliable?

What the research has found...

- **Replicates:** Irrational, show similar JDM biases, few significant differences in effect sizes (Goodman et al., 2013; Paolacci et al., 2010)
 - Framing
 - Conjunction fallacy
 - Outcome bias
 - Present bias
 - Delay/expedite asymmetries
 - Certainty effect
 - Risk aversion
- More info:
 - See blog page “Experimental Turk / Resources”
 - Review article Paolacci & Chandler (2014)

CDPS

Generate Questions

- Find 3 or 4 people next to you
- Introduce yourself
- Write down 3 questions/issues with Mturk
- Crowdsource the answers

MTurk is blocked for non-US requesters. What do I do?



- Payments Accounts now require an SSN or an EIN
- <https://news.ycombinator.com/item?id=8208718>
- Have institution open an account with an EIN
<http://www.irs.gov/Businesses/Small-Businesses-&-Self-Employed/How-to-Apply-for-an-EIN>
- Mturk intermediaries: <http://mturkdata.com/academic-surveys.html>

Alternatives to MTurk (especially non-US)?



- <https://www.socialsci.com/>
- <http://www.crowdfunder.com/>
- <https://prolificacademic.co.uk/>
- <http://www.cognilab.com/>
- <http://clearvoiceresearch.com/>
- ...

Guidelines for academic researchers

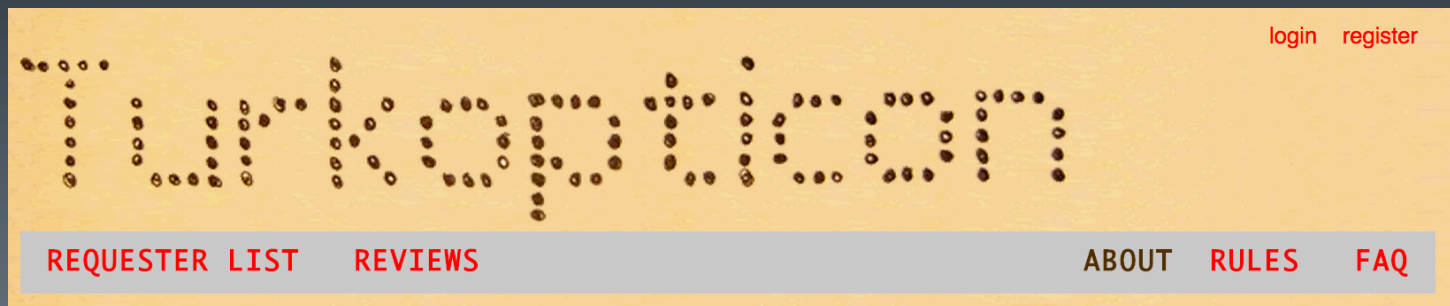


- Identify yourself (name, institution, lab, etc.), requester privacy is secondary
- Provide reasonable time estimates
- Approve work promptly (1-3 days, 7 max)
- Maintain worker privacy
- Abide by AMT terms of service
- Be clear when rejecting work
- Do not block workers
- Be responsive
- Pay fair

[http://wiki.wearedynamo.org/index.php/
Guidelines for Academic Requesters](http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters)

Compensation: What should I pay workers?

- No compensation effects found for common surveys (Buhrmester et al., 2011; Goodman et al. 2013)
- But pay fair!
- Keep in mind the Minimum Wage
 - \$0.92/hr in India (\$.015/minute)
 - \$7.25/hr in US (\$.12/minute)
 - \$9.32/hr in Washington state (\$.15/minute)
- Also keep in mind that you are not anonymous:



Do journals/editors accept it?

What can I do for them to accept it?

What issues do I need to be worried about?

- Any negative experiences?
- Solutions?
 - Provide details: dropout rate, gender, location, compensation, filters/qualifications
 - Cite papers on Mturk (see blog)
 - Explain why Mturk is good for your specific study
- Opposite problem: Editors requesting a quick Mturk survey to replicate a finding
 - Is every study good for Mturk?

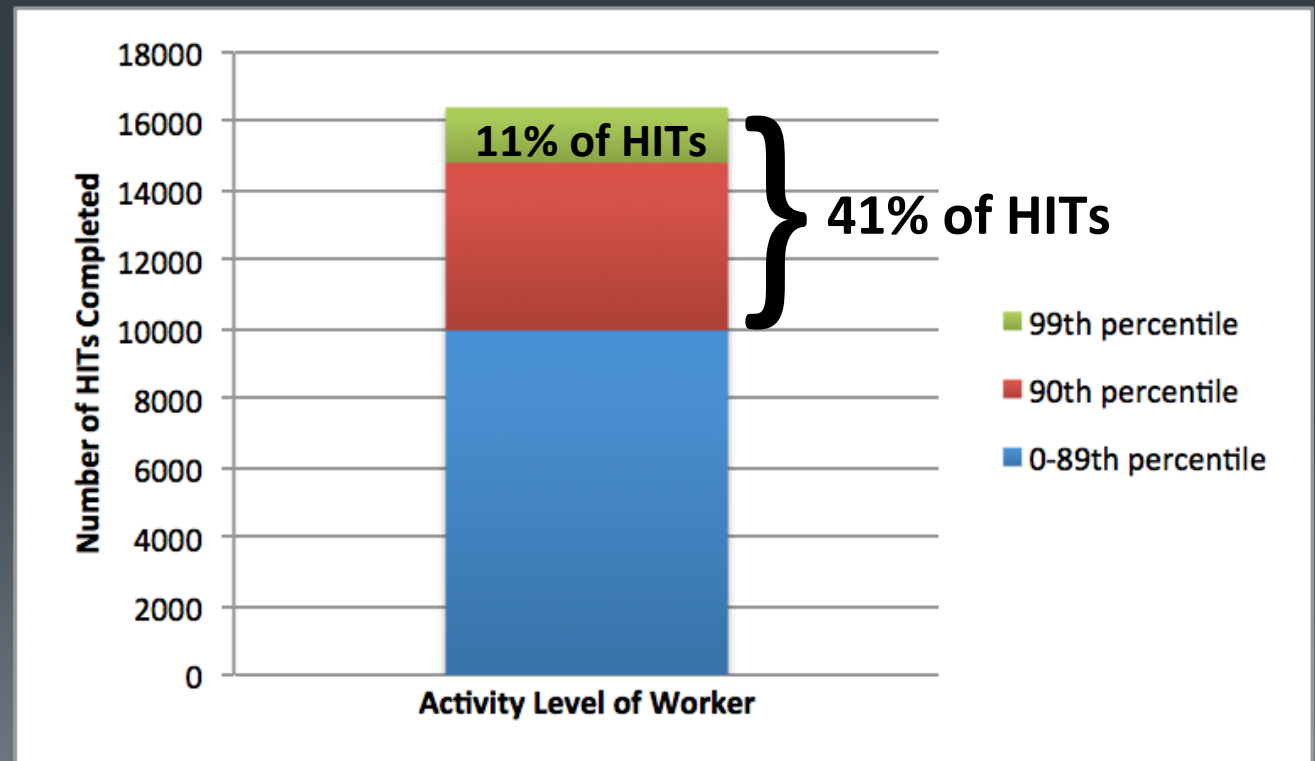
Is every study good for Mturk?

- Any negative experiences?
- Repeated participants (see Chandler, Mueller & Paolacci 2014 *Behavior Research Methods*)
- Common manipulations & paradigms
 - Power primes, CRT, etc.
- Some effects may be harder to show
 - Objective tasks with online answers
 - Valuations, or primes of \$ or time
 - Turkers are more tightwad w/ spending, more materialistic, lower value of time (Goodman et al., 2013)
 - Individual differences may need calibration
 - Turkers low on extroversion, emotional stability, self-esteem (Goodman et al., 2013)



Do I have to worry about repeat participants?

- Most HITs done by just a few turkers
- Chandler, Mueller, & Paolacci 2014 BRM:
10% of workers responsible for 41% of HITs!



How do I prevent repeat participants?



- Block previous workers
 - “Please don’t do this survey if you’ve done it already.”
 - Direct turkers to same page, change the redirect
 - “Have you done anything similar before?”
 - PsiTurk, TurkGate
 - Qualtrics/MTurk script on Experimental Turk blog
 - Exclude workers based on **qualifications (new!)**
 - **Assign Qualification** to Study 1 participants
 - Require Study 2 participants not to have that qualification
- <http://turkrequesters.blogspot.nl/2014/08/how-to-block-past-workers-from-doing.html>

How do I exclude using qualifications?

- Create a new Qualification

The screenshot displays a web application interface. The top navigation bar includes 'Home', 'Create', 'Manage', 'Developer', and 'Help'. The 'Manage' tab is active, and the 'Qualification Types' sub-tab is selected. The main content area shows 'Manage Qualification Types' with a message: 'Below is a list of your Qualification Types and the corresponding number of Workers assigned to each.' A yellow button labeled 'Create New Qualification Type' is visible. A yellow warning box states: 'You have not created any Qualification Types. Please use the button above to create a new Qualification Type.' A modal dialog titled 'Create New Qualification Type' is open, containing the following text: 'This will create a new Qualification Type that can be assigned to Workers.' The 'Friendly Name' field contains 'project X, study 1'. The 'Description' field contains 'Workers who participated in study X (October 2014)'. Below the fields, a note states: 'The Qualification Type you are creating does not require Workers to take a Qualification test. You can create a Qualification Type that requires Workers to take a test using the Developer Tools. [Get more info.](#)' At the bottom right of the modal, there are 'Create' and 'Cancel' buttons.

How do I exclude using qualifications?

- Assign qualification to workers who participated in the relevant study
- (You can automate this using **Command Line Tools**)

The screenshot displays the MTurk web interface. The top navigation bar includes 'Home', 'Create', 'Manage', 'Developer', and 'Help'. The 'Manage' tab is active, showing 'Results', 'Workers', and 'Qualification Types' sub-tabs. The main content area is titled 'Manage Batches > Review Results' and contains instructions for reviewing and approving results. A table of assignments is visible, with columns for 'HIT ID', 'Worker ID', and 'Lifetime'. An 'Assign Qualification Type' dialog box is overlaid on the interface, prompting the user to select qualification types for a specific worker. The dialog box includes the following text: 'Select which of your Qualification Types to assign to this Worker. You will be prompted to provide a score between 0 and 100. (Assign up to 5 Qualification Types at a time.)' and a list of qualification types, including 'project X, study 1'. The dialog box also features 'Assign' and 'Cancel' buttons.

Home Create Manage Developer Help

Results Workers Qualification Types Manage HITs individually

Manage Batches > Review Results

Review Results

Select the check boxes on the left to approve or reject results. You only pay for approved results.

For additional batch information, view batch details.

selling books

Customize View Filter Results

150 of 150 assignments (FILTER APPLIED: only show assignments that are in 'Approved' status)

HIT ID ▲	Worker ID	Lifetime
3NZ1E5QA6Z2VOPZNSJUXP2Q672Y5B0		
3NZ1E5QA6Z2VOPZNSJUXP2Q672Y5B0		
3NZ1E5QA6Z2VOPZNSJUXP2Q672Y5B0		
3NZ1E5QA6Z2VOPZNSJUXP2Q672Y5B0		

Worker ID: A3F0NFNL0LFJT0

Worker Status for your work Never Blocked

APPROVAL RATE ON YOUR ASSIGNMENTS

Lifetime	100% (4/4)
Last 30 days	100% (2/2)
Last 7 days	0% (0/0)

WORKER ACTIONS:

Bonus Worker

Block Worker

Assign Qualification Type

Assign Qualification Type

Select which of your Qualification Types to assign to this Worker. You will be prompted to provide a score between 0 and 100. (Assign up to 5 Qualification Types at a time.)

project X, study 1

Assign or Cancel

How do I exclude using qualifications?

- When you run the new study, tell Mturk you don't want people with the qualification

Advanced Worker requirements «

Worker requirements:

Customize Worker Requirements...

Specify ALL the qualifications Workers must meet to work on your HITs:

HIT Approval Rate (%) for all [remove](#)

project X, study 1 [remove](#)

(up to 5)


Only Workers who qualify to do my HITs can preview my HITs.

Yes No

Incentive-compatible designs: Can I do it? Is it enough for econ journals?

- Yes! Incentivize using bonuses (we'll talk more)
- Econ journals
 - Mturk is not a “Clean Lab”: “Deception” used, so some economists may question it
 - Searched for articles in *QJE*, *AER*, & *Econometrica*:
 - Only one used Mturk (pilot study; Olea & Strzalecki 2014)
 - But it might change
 - See Horton, Rand, & Zeckhauser (2011) *Exp Economics*

How do I give a bonus? Can it be automated?



Multiple Options


- Use the web interface. Click on Worker ID, one worker at a time (or hire an RA to do this...)
- Use **Command Line Tools**
 - <https://aws.amazon.com/developertools/Amazon-Mechanical-Turk/694>
 - More complicated, but you can give to multiple workers at a time
 - See Chandler, Mueller, & Paolacci (2014) *BRM*
- Use the MTurk API

Is Mturk suitable for long-ish studies?



- One “validation” paper (not available yet)
 - Kleinberg, Mona, Richard R. Lau, Tessa Ditonto, and David J. Andersen, “Testing the Validity of Mechanical Turk Data Beyond Simple One-Shot Experiments.”
- Other examples?
 - E.g., Mor, S., Morris, M. W., & Joh. J. (2013). Identifying and Training Adaptive Cross-Cultural Management Skills: The Crucial Role of Cultural Metacognition. *Academy of Management Learning & Education*. (about 30 minutes)

Is Mturk suitable for longitudinal studies? How do I do it?



- Yes – redemption is high (e.g., 50% after one year)
- Typically:
 - Conduct Wave 1
 - Assign Qualification to Wave 1 participants
 - Require Qualification for participation in Wave 2
 - Message participants via Mturk, using either
 - Automatic scripts (e.g., one on the blog)
 - Hire an RA to do it manually via Mturk
 - Don't ask for email (against Mturk rules)

More questions



- My school doesn't allow Mturk. What do I do?
 - Usually an issue with accounting
 - Anyone overcome the issue?
- My advisor won't let me use Mturk, should I find a new advisor?

Want more info? Read these papers...



- Paolacci, Chandler, & Ipeirotis (2010) *Judgment and Decision Making*
- Goodman, Cryder, & Cheema (2013) *Journal of Behavioral Decision Making*
- Buhrmester, Kwang, & Gosling (2011) *Perspectives in Psychological Sciences*
- Horton, Rand, & Zeckhauser (2012) *Experimental Economics*
- Crump, Donnell, & Gureckis (2013) *PLoS ONE*
- Paolacci & Chandler (2014) *Current Directions in Psychological Science*
- See the blog <http://experimentalturk.wordpress.com/>