Running Studies Online

JOSEPH K. GOODMAN

The Ohio State University
Fisher College of Business

For more info or citations:

Special Thanks to my Coauthors

GABRIELE PAOLACCI
ERASMUS UNIVERSITY

SCOTT WRIGHT
PROVIDENCE COLLEGE
What are your questions about Mturk?
Mturk Myth or Truth?

1. Not real people
2. Low quality data
3. Uneducated
4. Poor/low income
5. Internet geeks
6. The lab is still more reliable
7. They don’t pay attention
8. Mostly Indian
9. Mostly non-native English speakers
10. It is Mturk not MTurk, or mturk
11. Editors/reviewers hate it!
   ◦ Won’t accept papers with only Mturk data
   ◦ And other views about Mturk
12. Bots! Bots! Bots!
13. Non-naivete
14. Drop-outs
15. Pre-screening
What Do Researchers Think about Mturk?

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Neither agree nor disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papers should not have all their data from Mturk</td>
<td>4.68</td>
<td>4.68</td>
<td>4.72</td>
</tr>
<tr>
<td>Mturk studies should be re-run using a non-Mturk sample.</td>
<td>3.64</td>
<td>3.15</td>
<td>3.76</td>
</tr>
<tr>
<td>Mturk samples are better than student samples.</td>
<td>4.00</td>
<td>3.96</td>
<td>4.04</td>
</tr>
</tbody>
</table>

- Papers should not have all their data from Mturk
- Mturk studies should be re-run using a non-Mturk sample.
- Mturk samples are better than student samples.
AE’s & Editors are Not All Convinced

"Papers should not have all their data from Mturk"

Distribution

<table>
<thead>
<tr>
<th></th>
<th>All P's (including quant and CCT)</th>
<th>ERB</th>
<th>AEs/Editors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td>0.11</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>0.20</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>5</td>
<td>0.18</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>6</td>
<td>0.15</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>0.24</td>
<td>0.30</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Depends on Area/Methodology

- Papers should not have all their data from Mturk
- I question the validity of data obtained using Mturk
Researcher Trust in Mturk

Trust seems to be correlated with use and experience
- Endogeneity issue: hard to know what causes what

Researchers expect Mturk use to grow

As use grows, trust may also grow

Other solutions? (other than waiting 10 years for more trust)
Solution 1: Stop Doing MTurk Studies
Solution 2: Get a job at THE Ohio State University

(OR ANOTHER SCHOOL WITH 8,000 SUBJECT HOURS/YEAR)
Solution 3: Other Platforms?

Prolific
- Requires > $6.50/hr
- Representative samples (extra $)
- Pre-screening (not extra $)
- Slightly slower
- Simple and more integrated
- Slightly less attention?

CrowdFlower

Qualtrics Panels

Positly and ClourdResearch/TurkPrime

Others?

Peer, Samat, Brandimarte, & Acquisti (2017) *JESP*

# of failed attention check questions between the platforms (% of P’s)

<table>
<thead>
<tr>
<th></th>
<th>MTurk</th>
<th>CrowdFlower</th>
<th>Prolific Academic</th>
<th>CBDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (%)</td>
<td>5%</td>
<td>18%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>1 (%)</td>
<td>9%</td>
<td>12%</td>
<td>19%</td>
<td>22%</td>
</tr>
<tr>
<td>2 (%)</td>
<td>27%</td>
<td>25%</td>
<td>22%</td>
<td>20%</td>
</tr>
<tr>
<td>3 (%)</td>
<td>57%</td>
<td>18%</td>
<td>52%</td>
<td>46%</td>
</tr>
<tr>
<td>4 (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Solution 4: Measure Real Behavior on MTurk

Ex₁: Tonietto, Malkoc, & Nowlis (2019)
- P’s entered their availability/calendar on Qualtrics
- IV: Generate times bounded/not bounded by tasks (btw SS)
- DV: Choice of MTurk study
  - 30 minute brief study for $2.50
  - 45 minute full study for $5.00
  - Neither

Ex₂: Coin flip game to measure cheating (Balasubramanian, Bennett, and Pierce 2017)
Solution 4: Measure Real Behavior on MTurk

Ex₃: Take picture of artwork and upload or email it

Ex₄: Costello and Malkoc (working paper)
  ◦ DV: Donate $ or donate time (micro-donations: counting/classifying animals, algae, weather patterns)

  ◦ Do you have other examples?
Solution 5: Go Beyond Surveys

Longitudinal studies
- 60% retention rates (but need to pay more)

Group interactions

Webcams for facial expression analysis and analyzing emotions (FaceReader)

Physiological measures
- Measure heart rate (Muender et al. 2016)

RA Work
- classification of content
- web searches
- registering for websites
- viewing ads
- edit transcripts
- test software
- skilled tasks: logo & graphic design, ad copy
  - See Fiverr, Upwork, Guru
Issue: Data Quality and Bots

How bad is it?
- Different opinions, and changed over time
- Bot-gate: August 2018
  - Evidence that they are foreign workers with low English proficiency, posing as US workers

SOLUTIONS

Exclude data points afterwards? NO!
Attention checks? Maybe. Be skeptical...
- Can facilitate p-hacking and researcher degrees of freedom
- Must be objective and pre-registered
- Cat and mouse game

Increase your sample size: instead of n=400 run 500
Reject bad data? Will Upset workers
Use TurkPrime or Positly
TaskMaster: cut and paste html into qualtrics

Issue: Pre-screening & Non-naiveté

Non-naiveté

◦ We may be sampling the same people across studies
◦ Sampling same 7,300 workers on average (Stewart et al. 2015)
◦ Consider sampling workers with less experience (eg, < 500 HITs)

Workers are usually honest, but will lie to be eligible for participation

◦ 48/106 heterosexual workers claimed LGB to be eligible for study
◦ 8% women, 17% men changed gender to be eligible for study

Lying increases with compensation

Lie more about owning a dog than religion or politics

Chandler & Paolacci 2017, Wessling, Huber, & Netzer 2017
A Solution to Many Problems: CloudResearch (website formerly known as TurkPrime) & Positly

- Pre-Screen based on demographics or other criteria (e.g., gender, political views, etc.)
- Non-naiveté: Block repeat participants
- Calculate Completion & Bounce rates
- Provide real random completion codes and automatically approve
- MicroBatch & HyperBatch
- Block duplicate IPs
- Anonymize worker IDs

Positly: Another solution

Limit participants using high quality checks:
- Block duplicate IP addresses
- Block suspicious IP addresses
- Require high approval rate and experience
- Require attentive participants
- Restrict IP address geolocation by Country
Pre-screening on TurkPrime: Two Ways

1. Pay to use TurkPrime’s criteria
   ◦ Expensive, but lots of options and fast

2. Create your own database (see supplemental slides)
   ◦ Stage 1: Collect important info (eg, gender, age, political affiliation)
   ◦ Stage 2: Launch HIT limited to target group
Guidelines for Researchers

Use qualifications and pre-screening
- > 95% approval
- Do not let workers to screen themselves

Be Ethical – pay a fair wage
- Min wage of $7.25/hr in US ($0.121/min)
- Paying more ≠ better?

Collect Data and Track your participants
- Collect basic demographics
- Prevents non-naiveté, anonymous via TurkPrime

Try other sources
- Attention checks: Keep objective and pre-register. Or increase sample size

Guidelines for Academic Requesters

- Clearly identify yourself
- Provide reasonable time estimates
- Approve work as soon as possible
- Maintain worker privacy
- Do not block workers
- Maintain a responsive line of communication

Monitor your reputation.
- Reddit forms
- Turkopticon (shut down)
- Others?

http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters
Other Questions?

NEXT UP: HOW TO CREATE YOUR OWN DATABASE...
How to Create Your Own Database of Workers
Many Issues are Solved by TurkPrime

See our paper "TurkPrime.com: A versatile crowdsourcing data ..." in Behavior Research Methods

Recruit Online Participants

Simple • Affordable • Powerful Crowd|Research

Toolkit for Amazon Mechanical Turk
- Basic Toolkit Free for Academic Users
- MTurk Panels on Your Own Requester Account

Lab Service: Mechanical Turk Panels
- Target MTurk Workers by Ethnicity, Age and Custom Traits

Lab Service: Prime Panels
- Recruit from 20 Million Participants by any Demographic

Over 5000 Labs and Researchers Worldwide Use TurkPrime

QUINNIPIAC UNIVERSITY
Tulane University
UNIVERSITY OF WATERLOO
Université de Montréal
NANYANG TECHNOLOGICAL UNIVERSITY
THE OHIO STATE UNIVERSITY
FISHER COLLEGE OF BUSINESS
Create a Study

Mechanical Turk Studies

Run a study on your own MTurk account
- MTurk account required.
- Gives you the flexibility to set payment amounts and other custom settings.
- Directly communicate with workers and monitor your own study progress.

Run a fully managed MTurk study
- No MTurk account needed.
- TurkPrime manages worker communication and study maintenance.
- Reach your desired population with our TurkPrime guarantee.

Need to run a complex study? Click here for a price quote.
<table>
<thead>
<tr>
<th>Question</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic Disability</td>
<td>Do you have a disability or chronic illness of any type (either physical or mental)?</td>
</tr>
<tr>
<td>Migraines</td>
<td>How often do you experience migraines, if at all?</td>
</tr>
<tr>
<td>Future Exercise</td>
<td>Do you intend to engage in regular exercise in the next six weeks?</td>
</tr>
<tr>
<td>Dieting</td>
<td>How often are you dieting?</td>
</tr>
<tr>
<td>Fitness Tracker</td>
<td>Do you regularly use a pedometer or any other type of fitness tracker?</td>
</tr>
<tr>
<td>Chronic Pain</td>
<td>Do you have a diagnosis of chronic pain (pain lasting more than 12 weeks)?</td>
</tr>
<tr>
<td>Occupation</td>
<td>Which best describes your current occupation/job/career; if student, career goal?</td>
</tr>
<tr>
<td>Current Status</td>
<td>Which of the following best describes your current status?</td>
</tr>
<tr>
<td>Industry</td>
<td>Please select the industry that best describes your work and/or field of training.</td>
</tr>
<tr>
<td>Current College Student</td>
<td>Are you currently?</td>
</tr>
<tr>
<td>Supervisor</td>
<td>Are you employed full time in a supervisory position or higher where people report directly to yi</td>
</tr>
<tr>
<td>Hours Employed</td>
<td>How many hours per week do you work?</td>
</tr>
</tbody>
</table>
Title: Short Survey

Title of the survey that will be displayed to workers

Description: Answer a short survey. Takes about 1 minute!

A description of your survey

Custom Instructions:

Survey takes about 1 minute.

Note: This HIT is an academic study being conducted by university researchers. Your answers are completely anonymous and we are just interested in your opinions. It is not being conducted for commercial purposes or to make profit, and it won't be sold or used by a company or marketers. You don't have to search the internet on this HIT. It's just opinion questions, so please take your time to answer the questions honestly.

Requirements:
- Must be 15 or older

(Optional) It is best to include instructions on your survey landing page. Nevertheless, you may include custom instructions to display to workers.

Standard instructions that workers should follow the hyperlink and submit a secret code (if set) are always included and will follow the custom instructions, if specified.

Keywords: survey, study, academic, preferences

Keywords associated with your HIT

- This project may contain potentially explicit or offensive content, for example, nudity
Running Studies Online: Mechanical Turk, Issues, & Solutions

Worker Payment Per Survey
$ 0.1

Expected time to complete HIT
1 minutes

Time allotted per assignment
10 minutes

Hit expires in
7 days

Survey Hyperlink
https://osu.sr1.qualtrics.com/jfe/form/SV_3xa1WCRuJL9LSXq

Query String Parameters
Turk365 will add the following query string parameters to your URL so you can take advantage of them in your data collection. The parameters are: workerId, assignmentId, and hitId.
- workerId - The MTurk Worker ID
- assignmentId - A unique ID that Amazon generates for a worker taking your HIT
- hitId - The ID of your HIT

For example, if your Survey Link is https://www.example.com, we will add the above parameters to the URL, so the final URL might look like https://www.example.com?workerId=ABCD14356hitId=35hgfbdasassignmentId=ytwjebwufg. You can then take advantage of these parameters by setting up Qualtrics Embedded Data.

Scheduled Launch Time (UTC)
Please note that you may leave the Scheduled Launch Time blank if you want to launch your survey right away. This feature is useful if you want to automatically launch your survey at a later time.

Scheduled Launch Time (UTC)

Required Format UTC Time Zone
UTC Time Now: 02/14/2018 04:52 AM

Automatic Worker Bonus
$ 0.0
The amount of time you want your HIT to be active for

Survey Hyperlink

https://osu.az1.qualtrics.com/jfe/form/SV_3xa1WCRuJL9LSXh

Preview link in new tab

The URL of your survey

Query String Parameters

TurkPrime will add the following query string parameters to your URL so you can take advantage of them in your data collection. The parameters are: workerId, assignmentId, and hitId.

- workerId - The MTurk Worker ID
- assignmentId - A unique ID that Amazon generates for a worker taking your HIT
- hitId - The ID of your HIT

For example, if your Survey Link is https://www.example.com, we will add the above parameters to the URL, so the final URL might look like https://www.example.com?workerId=ABCD1435&hitId=35hgsfbdlu&assignmentId=X.

Take advantage of these parameters by setting up Qualtrics Embedded Data

Scheduled Launch Time (UTC)

Please note that you may leave the Scheduled Launch Time blank if you want to launch your survey right away. This feature is useful if you want to automatically launch your Survey at a later time.

Scheduled Launch Time (UTC)  

Pick an optional launch-off date...

Please note the time needs to be in UTC format so if you want to launch on Feb 15, 2018, in Eastern Time, the time should be 02/15/2018 07:00 PM

Automatic Worker Bonus

$
Now go to your Qualtrics Survey ➔ “Survey Flow” ➔ “Set Embedded Data” ➔ enter variable names (case sensitive!)
Running Studies Online: Mechanical Turk, Issues, & Solutions
Dynamic Completion Code For Qualtrics

Each worker will have a unique secret code. Your HIT can either be auto-approved or manually approved.

Setup for Qualtrics

1. Create or Edit a Survey
2. Click on "Survey Options"

Survey Options

3. In the Survey Termination Section: Click on Custom end of survey message...

Survey Termination

- Default end of survey message.
- Custom end of survey message...

4. If you have a Secret Key option that you previously created Select it and you are done/Otherwise, Click New Message
5. Enter Secret Key in the Description field. (Next time you can select this message.)
6. Click the Source
   button
7. Paste the following text exactly into the window as shown

<iframe src="https://www.turkprime.com/TurkPrime/TakeLaunchedSurvey/DynamicKey" width='100%' height='200'></iframe>

Create a New Message

- Category: End of Survey Messages
- Description: Secret Key

8. Click Save

Running Studies Online: Mechanical Turk, Issues, & Solutions
Auto Pay Workers
(approve only if completion code is correct)
Other Qualifications...

Worker Qualifications

Worker Level

- Any Level
- Master

Location

- is one of
- Select Locations (up to 30)
- UNITED STATES

Note Regarding HIT Approval Rate
When you set HIT Approval Rate (e.g., 80-100%) TurkPrime automatically sets Number of HITs Approved to be at least 100 since MTurk assigns an Approval Rate of 100% to Workers who have completed less than 100 HITs (MTurk API Reference).

HIT Approval Rate (%) for all Requesters' HITs

Number of HITs Approved

- 100
- 1,000,000

Worker needs to have completed between 100-1,000,000 HITs

Group Requirements

Create a new Worker Group here

- New Nalvete (Exclude most active Workers) (more details)

Nalvete is free until April 1, 2018.
MicroBatch HITs

- Please be advised that MicroBatch HITs take longer to complete. The average study takes 1 hour per 7 workers, but times may vary.
- MicroBatch will not be enabled if the Time allotted for your HIT is more than 90 minutes.
- Additionally, MicroBatch will not be enabled if you include Workers.

MicroBatch

The survey will be batched into small HITs. (Duration for survey must be 90 minutes or less and you cannot include any Workers for this to be enabled)

Auto Restart Time In Minutes

The time delay between successive MicroBatch HITs. If left blank, the value will be system optimized.

Number Participants Per Batch

The number of participants per MicroBatch HIT. Value of 9 or less will save 50% off MTurk fees.