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.wordpress.com/

Joseph K. Goodman & Gabriele Paolacci
- "HITs" (Human Intelligence Tasks) tasks for money ($0.01 - $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)

Joseph K. Goodman & Gabriele Paolacci

#mturkacr
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

<table>
<thead>
<tr>
<th></th>
<th>Previous Research</th>
<th>Monday’s Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>Male</td>
<td>58%</td>
<td>62%</td>
</tr>
<tr>
<td>ESL</td>
<td>56%</td>
<td>6%</td>
</tr>
<tr>
<td>US</td>
<td>47%</td>
<td>93%</td>
</tr>
<tr>
<td>India</td>
<td>33%</td>
<td>6%</td>
</tr>
<tr>
<td>Student</td>
<td>24%</td>
<td>21%</td>
</tr>
<tr>
<td>Some Univ</td>
<td>92%</td>
<td>94%</td>
</tr>
<tr>
<td>Bach Degree or higher</td>
<td>49%</td>
<td></td>
</tr>
<tr>
<td>Median Income</td>
<td>$35,000</td>
<td></td>
</tr>
<tr>
<td>Employed Full-Time</td>
<td>48%</td>
<td></td>
</tr>
<tr>
<td>Employed Part-Time</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>Ave # HITs per day</td>
<td>36.6</td>
<td></td>
</tr>
<tr>
<td>Mean # yrs on Mturk</td>
<td>1.37</td>
<td></td>
</tr>
</tbody>
</table>
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
• Crowdsourced 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](https://news.ycombinator.com/item?id=8208718)
- Have institution open an account with an EIN
- Mturk intermediaries: [http://mturkdata.com/academic-surveys.html](http://mturkdata.com/academic-surveys.html)
Alternatives to MTurk (especially non-US)?

- https://www.socialsci.com/
- http://www.crowdflower.com/
- https://prolificacademic.co.uk/
- http://www.cognilab.com/
- http://clearvoiceresearch.com/
- ...

Joseph K. Goodman & Gabriele Paolacci
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

Joseph K. Goodman & Gabriele Paolacci

#mturkacr
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 ($0.015/minute)
  • $7.25/hr in US ($0.12/minute)
  • $9.32/hr in Washington state ($0.15/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

How do I exclude using qualifications?

- Create a new Qualification
How do I exclude using qualifications?

- Assign qualification to workers who participated in the relevant study
- (You can automate this using Command Line Tools)
How do I exclude using qualifications?

• When you run the new study, tell Mturk you don’t want people with the qualification
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
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

Joseph K. Goodman & Gabriele Paolacci
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?
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*
- Crump, Donnell, & Gureckis (2013) *PLoS ONE*
- Paolacci & Chandler (2014) *Current Directions in Psychological Science*
- See the blog [http://experimentalturk.wordpress.com/](http://experimentalturk.wordpress.com/)