Using Amazon's Mechanical Turk™ to Recruit Transition Age Adults with Autism: A Cautionary Tale

Jordan A. Findley (MS)1, John McGrew (PhD)2 and Lisa A. Ruble (PhD)1
University of Kentucky1, Indiana University-Purdue University Indianapolis2

What is This Study About?

- Amazon’s Mechanical Turk™ (MTurk) is comparable to other online platforms and in-person laboratory studies within the social sciences (Thomas & Clifford, 2017).
- Researchers recently reported a decline in data quality on MTurk due Virtual Private Servers (VPS) allowing participants to circumvent traditional screening methods (Dennis, Goodson, & Pearson, 2018).
- Use of MTurk in clinical populations has led to serious concerns of individuals misrepresenting personal characteristics to meet eligibility requirements which could lead researchers to draw incorrect conclusions (Chandler & Paolacci, 2017; Kan & Drummey, 2018).

Research Questions

- R1. What percentage of participant attempts in a survey of daily living activities are usable in a clinical sample of individuals or their caregivers with a self-reported ASD diagnosis aged 18-22?
- R2. What are the most common reasons for non-payment (i.e., failed to meet study eligibility)?
- R3. What are qualitative and quantitative differences in responses from VPS-respondents and non-VPS respondents?

Method

Online crowdsourcing marketplace.

Anonymously pay participants for completing surveys.

Participats

- R1-R2: All based on self-report data
- R3: Based on pilot caregiver report data (after inclusion of Fake ASD scale and additional open ended responses)

Survey Study Design

- Linked Qualtrics survey to MTurk
- Set two Worker qualifications in MTurk as well as 13 Screen- out questions in Qualtrics
- VPS respondents were identified post-hoc by tracing IP addresses via https://www.iplocation.net/

Question Distribution

2 informed consent | 10 demographic | 33 community participation | 13 screening

Results

<table>
<thead>
<tr>
<th>Survey Responses</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Attempts</td>
<td>Total Unique Participants</td>
</tr>
<tr>
<td>% Usable = 6.5%</td>
<td>700</td>
</tr>
</tbody>
</table>

R1. Usable Responses

R2. Common Reasons for Non-Payment

- Over 50% of response attempts were deemed unusable within the first seven demographic questions because they did not meet inclusion criteria
  - Yes/No: Do you have a comorbid sensory disability?
  - Yes/No: Did you have an IEP in school?
  - Yes/No: Did you have a formal autism diagnosis?

R3. Quantitative Differences Between VPS and non-VPS Respondents

- On average, participants responding from VPS sources attempted the survey four times. Non-VPS participants attempted the survey two times
- Despite setting a qualification within MTurk to exclude international participants, some IP addresses were traced back to international sources.
- Rates of VPS respondents increased over time

Discussion

- Usable return rates were low with demographic deception occurring at high percentages.
- Research conclusions based on MTurk samples should be interpreted with caution as traditional screening methods may not be adequate for screening out low quality or invalid responses.

Recommendations

- Consistent with previous recommendations, it is important to incorporate open-ended responses within surveys to flag problematic responses (Dennis, Goodson, & Pearson, 2018).
- Incorporating programs such as Qualtrics that collect IP data is a useful post-hoc method to confirm participants are responding from non-VPS sources.
- Incorporation of multiple screen out questions, phrased in different ways can assist in identifying inconsistencies.

Note. Formatting was not edited.

Contact: For a full list of references, please contact Jordan Findley, jordan.findley@uky.edu