

# **How Well Do Attention Checks Identify Fraudulent Online Responses, and Which Types of Attention Checks are Best?**

**Emily Geisen, Stephen Poole**  
**Qualtrics**

**AAPOR 2022**  
**Chicago**

# Objectives

1. Assess whether attention checks are useful for identifying poor quality survey responses in online panels
2. Assess whether attention checks harm or improve respondents' survey behavior
3. Assess which attention checks tested work the best

# Methods and Sample

- Fielded in October 2021
- Study topic: Current events (politics, covid, health)
- Median speed: ~7 minutes
- 3,708 completes (duplicates removed)
- 5 online, nonprobability panels
  - Panel A: 788
  - Panel B: 778
  - Panel C: 726
  - Panel D: 655
  - Panel E: 761

# Rs randomly assigned to Attn Check Condition

Within each panel, randomly assigned to treatment condition

- **Control** (n=752)
- **Commit** to providing thoughtful answers (n=724)
- Write in the word “**Purple**” (n=734)
- **Speed** check with request to slow down (n=746)
- Which of these is a **vegetable**? (n=752)

# Attn Check: Commit

We care about the quality of our survey data. For us to get the most accurate measures of your opinions, it is important that you provide thoughtful answers to each question in this survey.

Do you commit to providing thoughtful answers to the questions in this survey?

I can't promise either way

Yes, I will

No, I will not

Had to pick Yes, I will

# Attn Check: Purple

The following question is to verify that you are a real person.

Please enter the word **purple** into the box below.

## **Attn Check: Speed**

You seem to have answered the first part of this survey very quickly. Please be sure you have given all the items in the question sufficient thought so that your answers are thoughtful and accurate.

# Attn Check: Vegetable

The following question is to verify that you are a real person.

Which of the following is a vegetable?

- Egg
- Salmon
- Broccoli
- Cheeseburger
- Pizza
- Milk



# Quality Indicators

Dupes or Fraud removed using digital fingerprinting technology

- 2 *other* quality indicators (besides attention checks):
  - Familiarity with a vaccine maker that does not exist
  - Being diagnosed with a disease that is very rare (<0.5% of population has)
  - Straight lining on 3 or more matrix questions (where SL would not be expected)
  - Speeding on 6 or more survey pages (out of ~40)
  - Inconsistent answers across questions (2 checks)
  - Trigger 2+ errors (e.g., age over 120)
  - Any bad open-ended (OE) question

# Results

# Percent of Rs failing each attention check

Attention Check	n Failing	Percent Failing
Control	NA	NA
Commit	34	5.2%
Purple	7	1.1%
Speed	29	4.3%
Vegetable	11	1.6%

# Of those who failed Attn Check, what % failed other Quality Indicators?

Condition	N	% failing 1+	% failing 2+	% failing open-ended
Commit	34	65%	44%	44%
Purple	7	100%	86%	100%
Speed	29	45%	24%	34%
Veg	11	91%	73%	55%

# Of those who passed Attn Check, what % failed other Quality Indicators?

Condition	n	% failing 1+	% failing 2+	% failing open ended
a. Control	690	21.6%	11.0%	13.7%
<b>b. Commit*</b>	<b>612</b>	<b>16.3%*</b>	<b>4.6%*</b>	<b>8.5%*</b>
c. Purple	656	19.7%	<b>7.5%*</b>	<b>10.1%*</b>
d. Speed	647	22.1%	9.7%	12.1%
<b>e. Veg*</b>	<b>678</b>	<b>16.4%*</b>	<b>6.7%*</b>	<b>8.9%*</b>

\* Treatment different from control for all measures at significance  $p < 0.05$

# Conclusions

- Rs who fail certain attention checks (Commit, Purple, Vegetable) are likely to fail other quality checks
- Seeing an attention check is associated with fewer quality issues later in the survey compared with not seeing attention check
- Multiple types of attention checks appear to be successful, but commit did marginally better than others at reducing poor quality responses later
- Certain demographic subgroups (e.g., lower education) are disproportionately affected

# Limitations / Future Research

- Limited to online, nonprobability panels
- Did not examine drop-off or feasibility rates
- Examine whether simply providing attention checks improves estimates compared with benchmarks
- If attention checks become more common (ubiquitous) will they be as effective?