

Disclaimer: Opinions expressed in this paper are those of the authors and do not reflect official policy of the U.S. Bureau of Labor Statistics.



Crowdsourcing in the Cognitive Interviewing Process

Joe Murphy¹, Jennifer Edgar², Michael Keating¹

¹ RTI International

² Bureau of Labor Statistics

jmurphy@rti.org
@joejohnmurphy
@surveypost

www.rti.org

Overview

- Traditional cognitive interviewing
- Crowdsourcing alternatives
- Study comparing traditional & crowdsourcing methods, topic: clothing expenditures (in part)
- Results
 - Recruitment
 - Comprehension
 - Response strategies
- Relative advantages and disadvantages
- Future research directions

Traditional cognitive interviewing

- Local participants recruited by:
 - Newspaper ads
 - Flyers
 - Word of mouth
 - Craigslist (Murphy et al., 2007)
- Conducted in person, 1 at a time, in a lab
- Think-aloud & follow-up probes to explore comprehension, retrieval, decision & response

Study design: traditional cognitive interviews

- 71 participants recruited via newspaper & online ads
- Screened on demographics to fill quotas
- DC, headquarters-based testing + 3 regional testing cities
- Interviewer administered questions, scripted & spontaneous follow up probes
- Audio recorded and interviewer notes
- Interviews lasted approximately 20 to 30 minutes

Crowdsourcing alternatives

- “Tapping into the collective intelligence of the public to complete a task.” (King, 2009)

- Distinctive features:
 - broad reach,
 - a motivated crowd,
 - participants well suited to complete the task,
 - infrastructure to facilitate the task completion.

- Many using crowdsourcing platforms for data collection (Keating et al., 2013)

Study design: crowdsourcing via TryMyUI

- Panel for remote website usability testing
- Developed quotas (e.g. 5 males with high school education) and submitted the task to TryMyUI
- Eligible participants sent task information & able to complete until quota filled
- 44 completed SurveyMonkey instrument, with cognitive follow-ups captured via audio
- TryMyUI limits tasks to 20 minutes; most completed in less time

Study design: crowdsourcing via Amazon Mechanical Turk

- Large base of workers (“Turkers”) ready to complete tasks
- Posted 4 separate tasks paying \$0.75 for 5 minute cognitive protocol and demographics, limited to U.S. 18+
- More than 250 participants per, taking only a couple days each to complete
- Web self-administered instrument in SurveyGizmo
- Reference questions above probes to aid respondents & prevent recall challenges

Study design: crowdsourcing via Facebook

- Tried 3 types of targeting:
 - 18+ U.S. English speaking (158M)
 - 18+ U.S. English speaking & “like” music (36M)
 - 18+ U.S. English speaking & “like” American Red Cross (1M)
- Ads promoted \$5 Amazon gift cards (with music image for type 2) and \$5 Red Cross donation
- Red Cross targeting by far the most effective (see Murphy, 2013 for more information)
- 60 interviews on SurveyGizmo over 2 weeks

Survey questions and probes same across modes

- Example:

Reference question: "Since the first of May have you or any member of your household purchased any swimsuits or warm-up or ski suits?"

3. What types of items did you think of when you read this question?

Characters used: 0 out of 500.

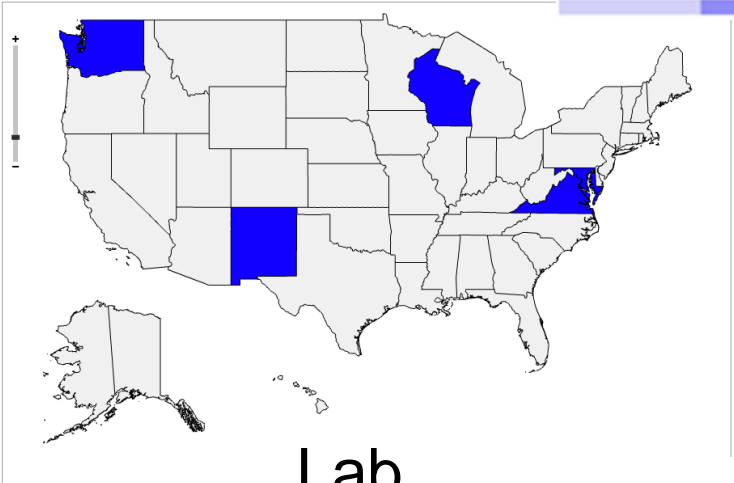
Back

Next

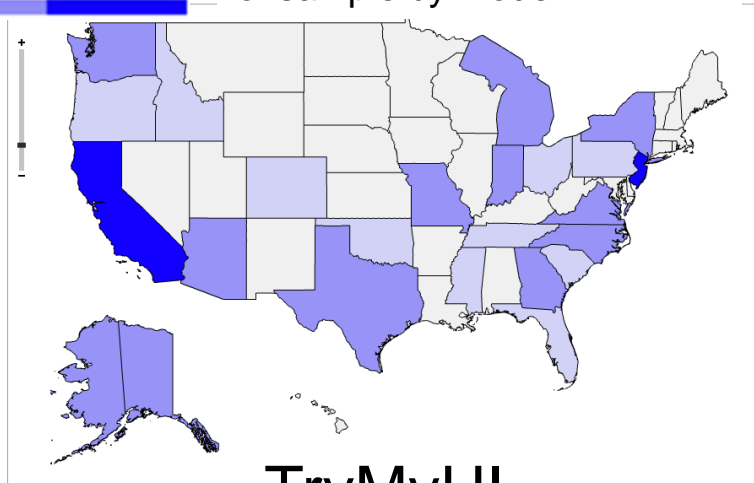
Results

Results: recruitment by location

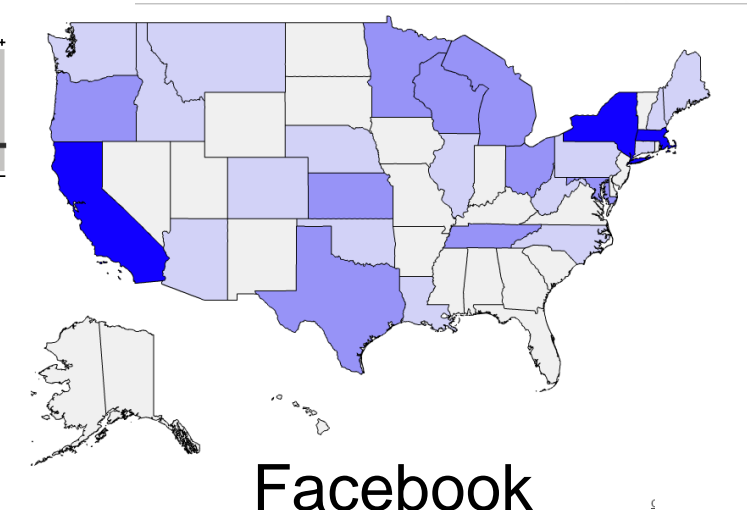
1% 5% 10%+ of sample by mode



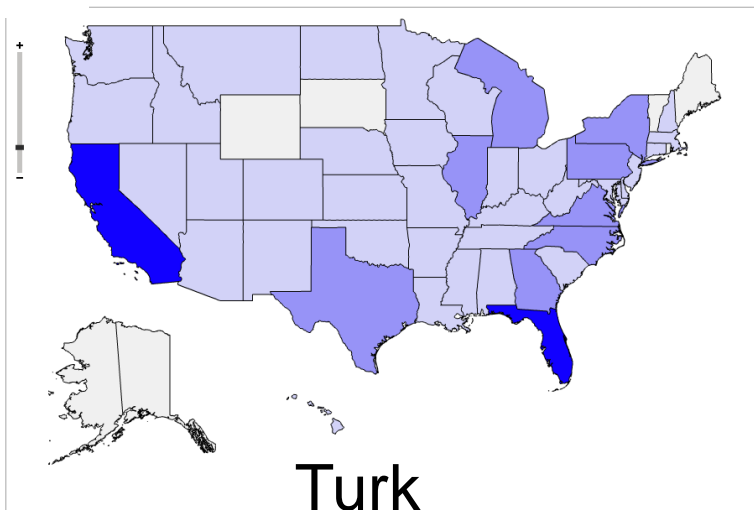
Lab



TryMyUI

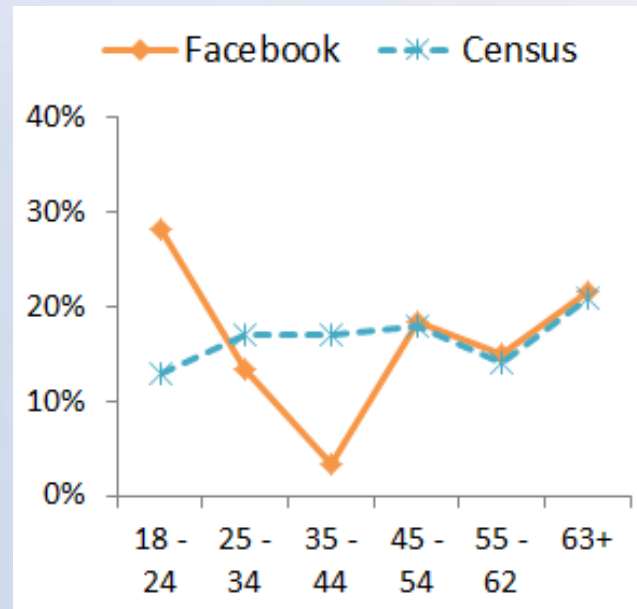
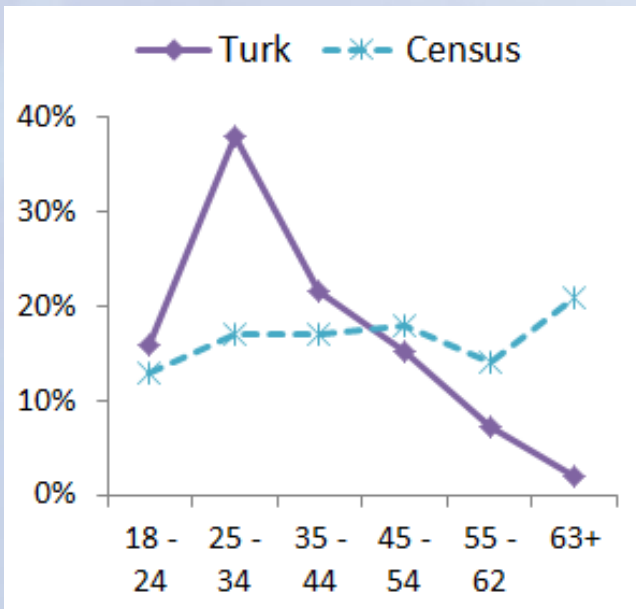
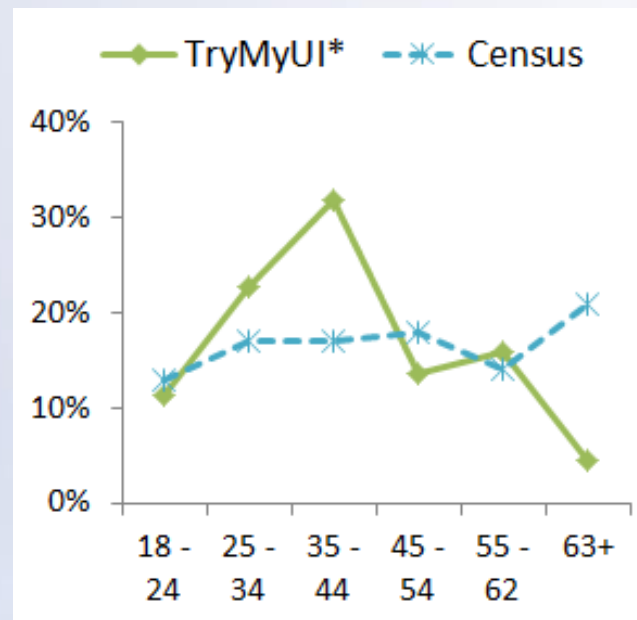
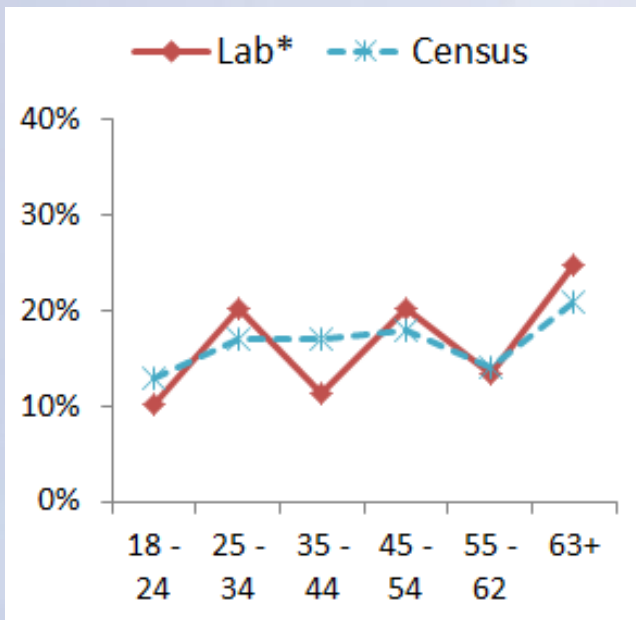


Facebook



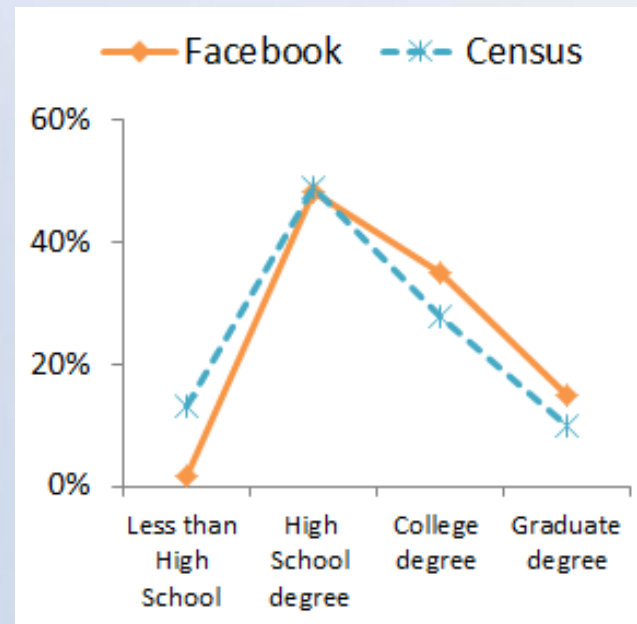
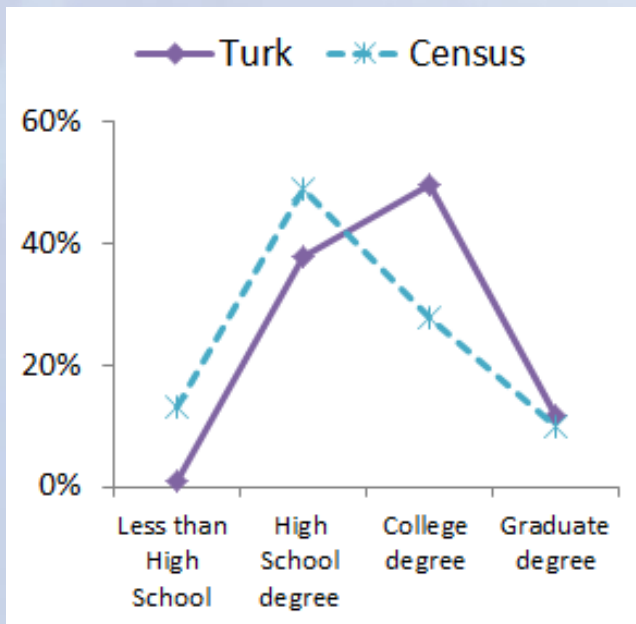
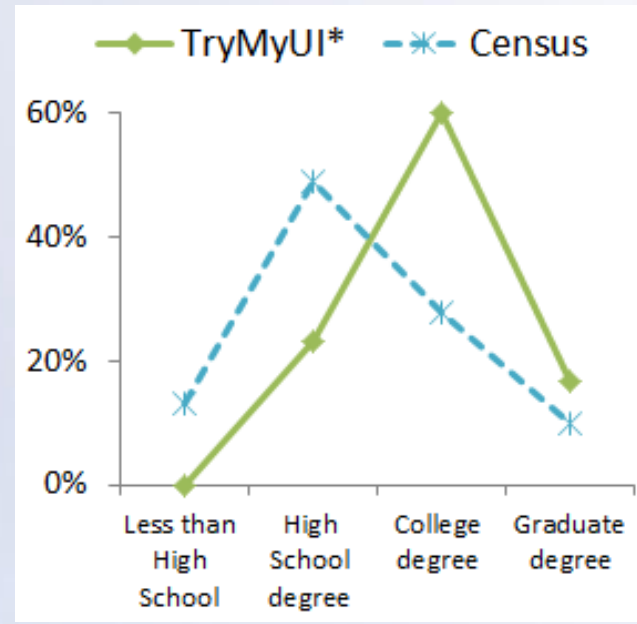
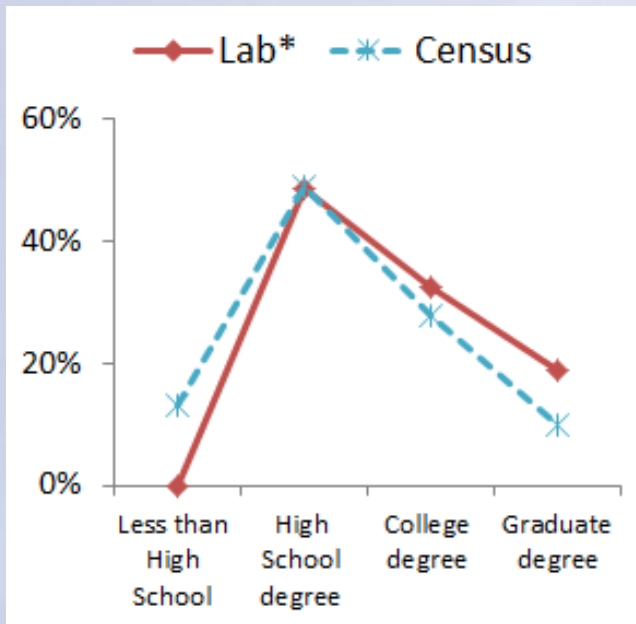
Turk

Results: recruitment by age



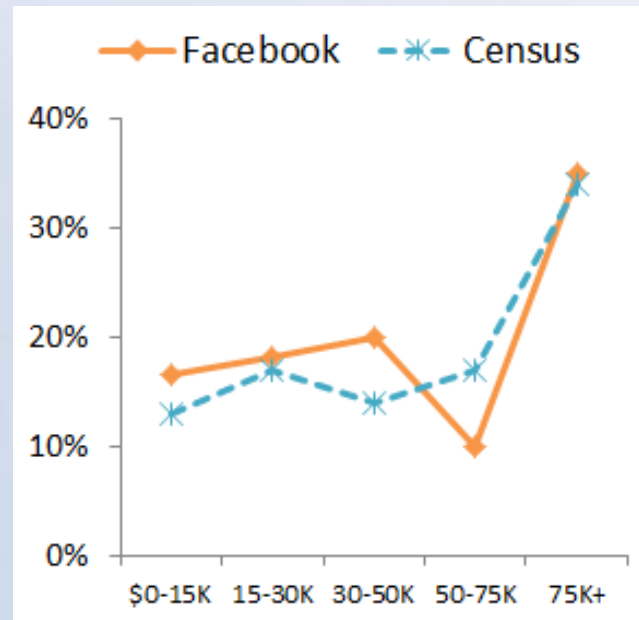
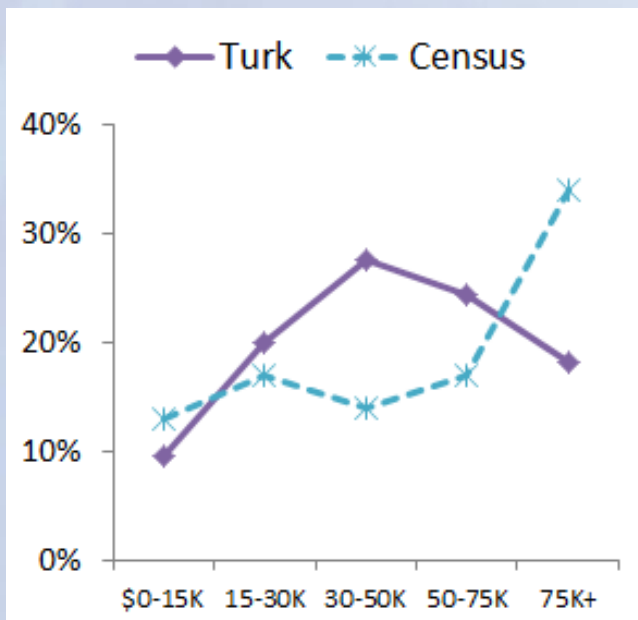
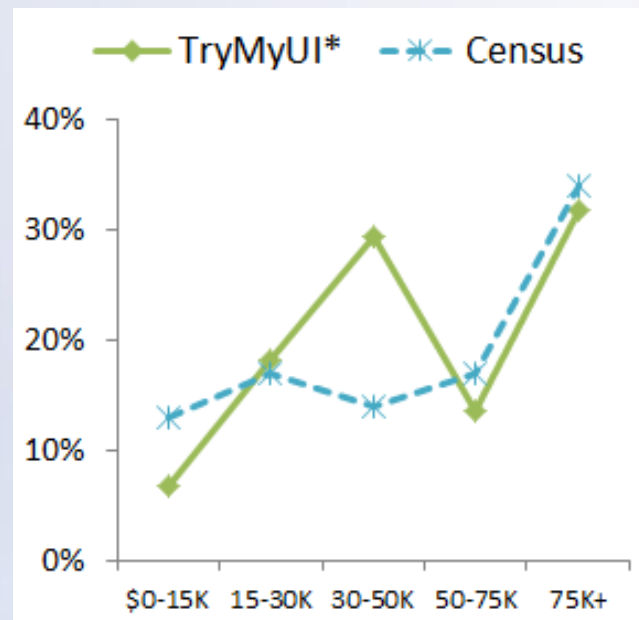
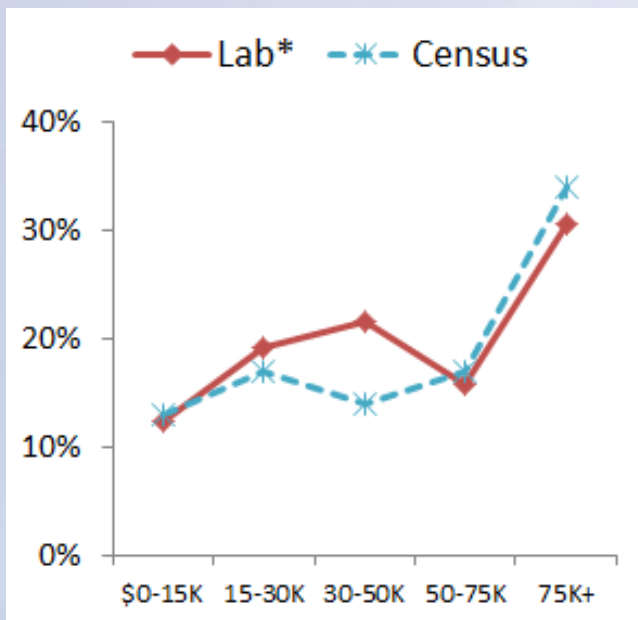
* Lab and TryMyUI recruitment used quota sampling

Results: recruitment by education



* Lab and TryMyUI recruitment used quota sampling

Results: recruitment by annual income

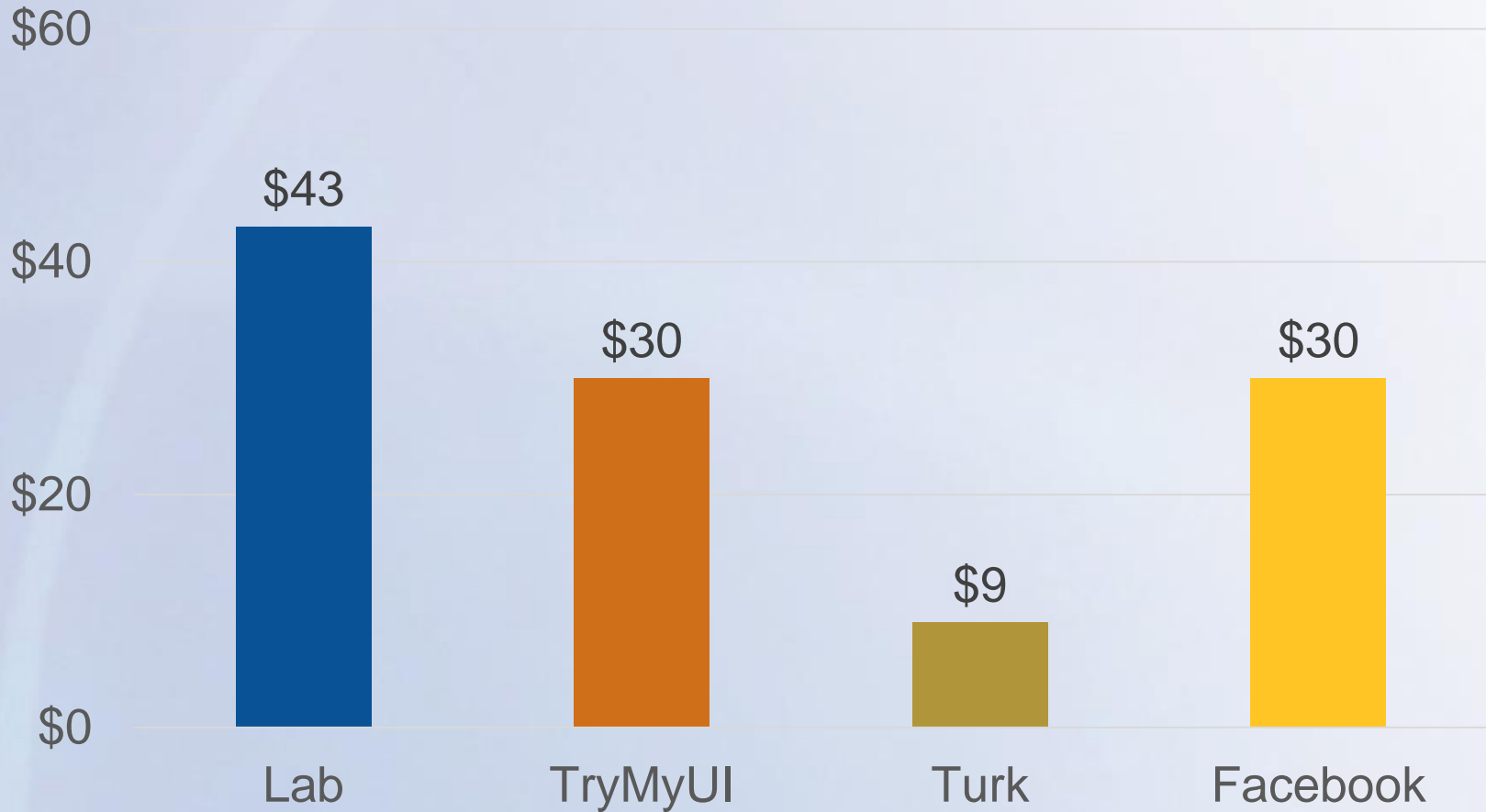


* Lab and TryMyUI recruitment used quota sampling

Results: participant characteristics summary

- The lab and TryMyUI recruiting used a quota method, so participants generally represented the US population in age and income
 - Even with the quota sampling, TryMyUI participants had higher levels of education than the US population
- Facebook and Turk did not use quota sampling, participants tended to be
 - Younger
 - More Educated (Turk)
 - Have slightly lower income

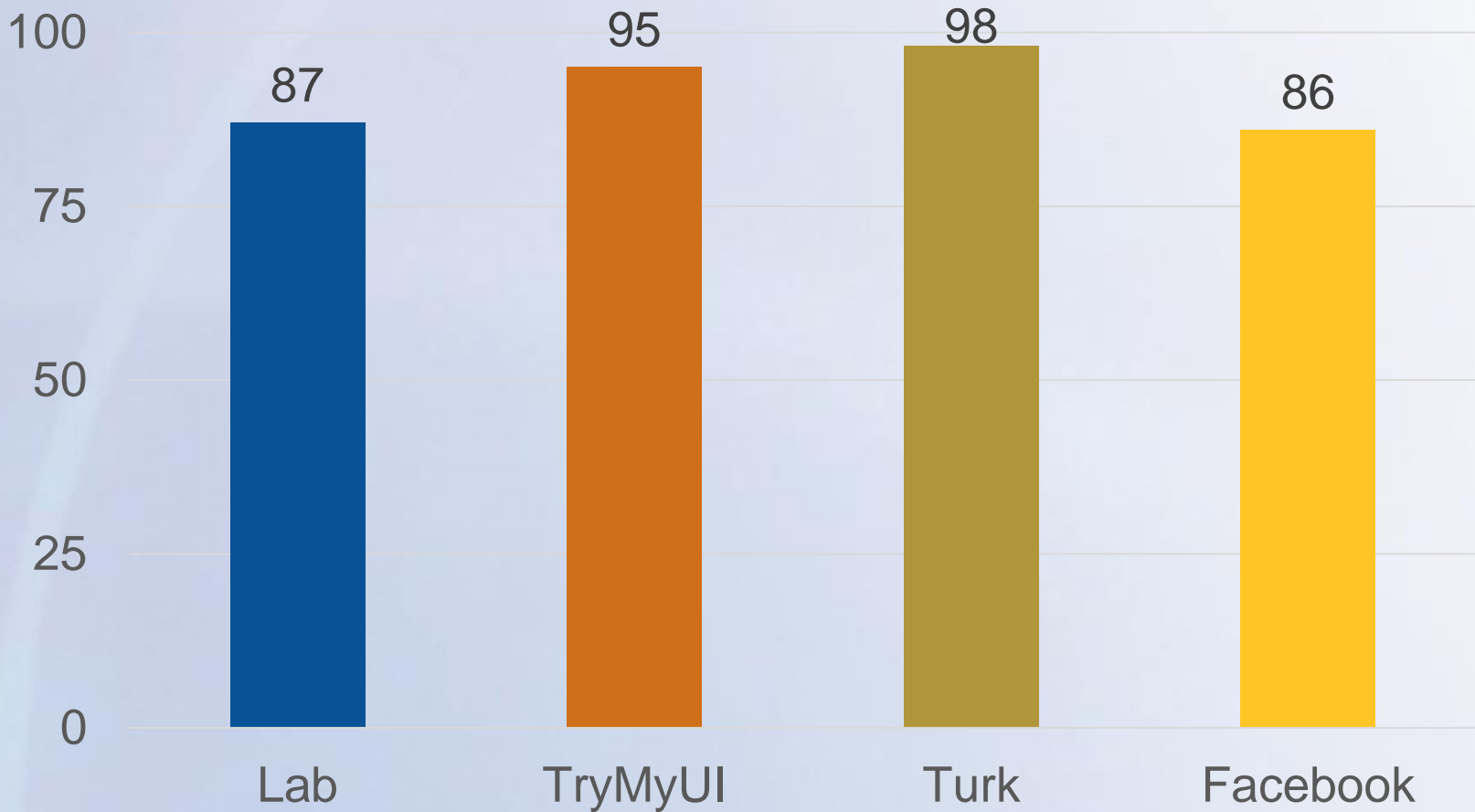
Results: incentive cost per hour



Results: comprehension

- Goal: understand participants' comprehension of expenditure question
- Participants asked: “Since the first of {reference month} have you or any member of your household purchased any swimsuits or warm-up or ski suits?”
- Follow-up: “What types of items did you think of when you heard the question?”

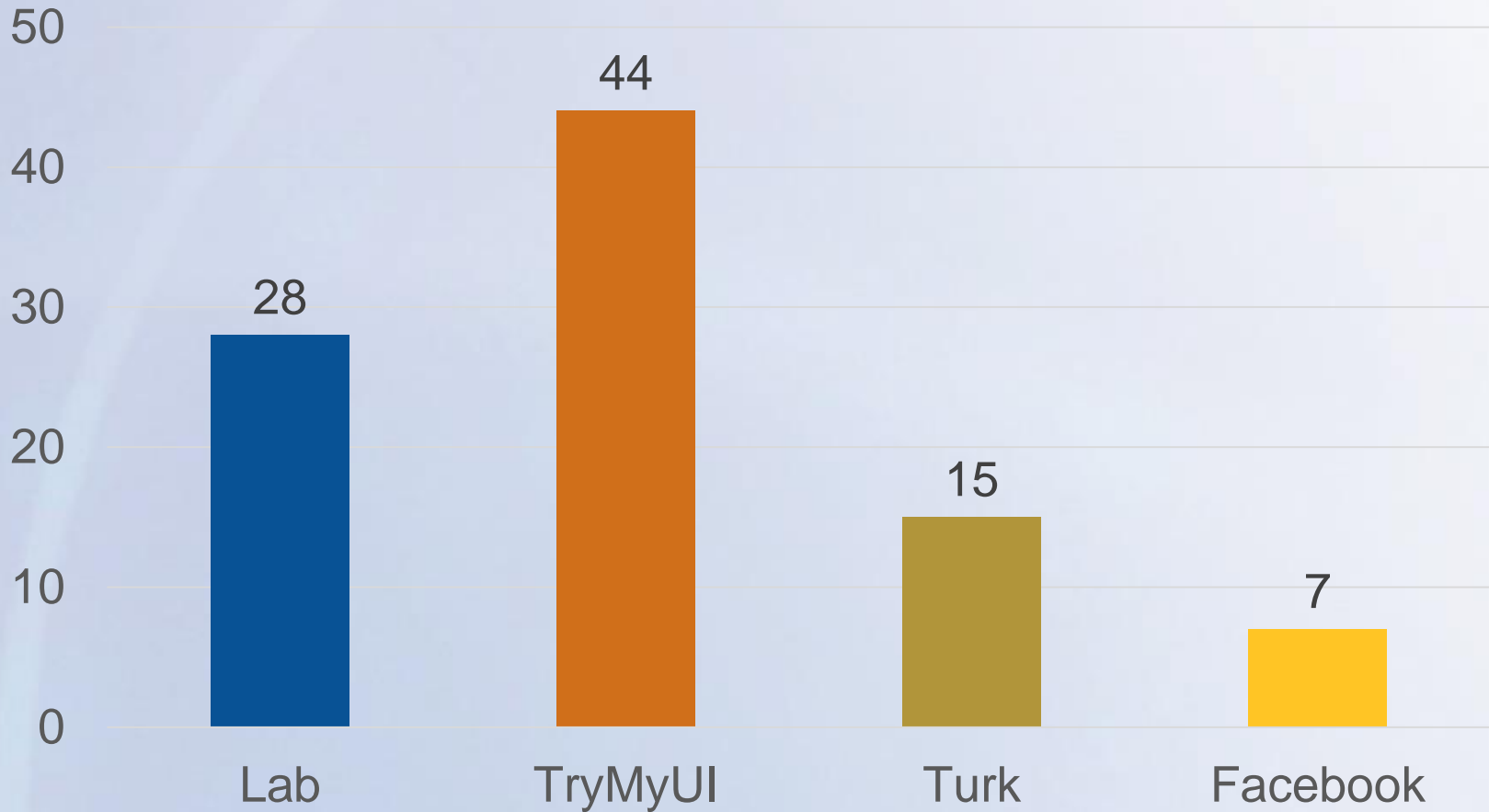
Results: comprehension, % relevant responses



Results: response strategy

- Goal: understand participants' comprehension of expenditure question
- Participants asked: “Since the first of {reference month} how much have you or any member of your household spent on clothing?”
- Follow-up: “How did you arrive at your answer?”

Results: response strategy word count



Results: response strategy quality

Each open ended response was coded for quality

Completely unusable	no information that could be used to code a response strategy
Some usable information	some information to identify a response strategy, but considerable probing would be needed to code a response strategy
Mostly complete	only a little probing would be required to be able to code response strategy.
Complete	enough information to be able to code response strategy without probing

Results: response strategy quality, examples

“Since the first of {reference month} how much have you or any member of your household spent on clothing?... *How did you arrive at your answer?*”

- **Completely unusable:** “Price”

Results: response strategy quality, examples

“Since the first of {reference month} how much have you or any member of your household spent on clothing?... *How did you arrive at your answer?*”

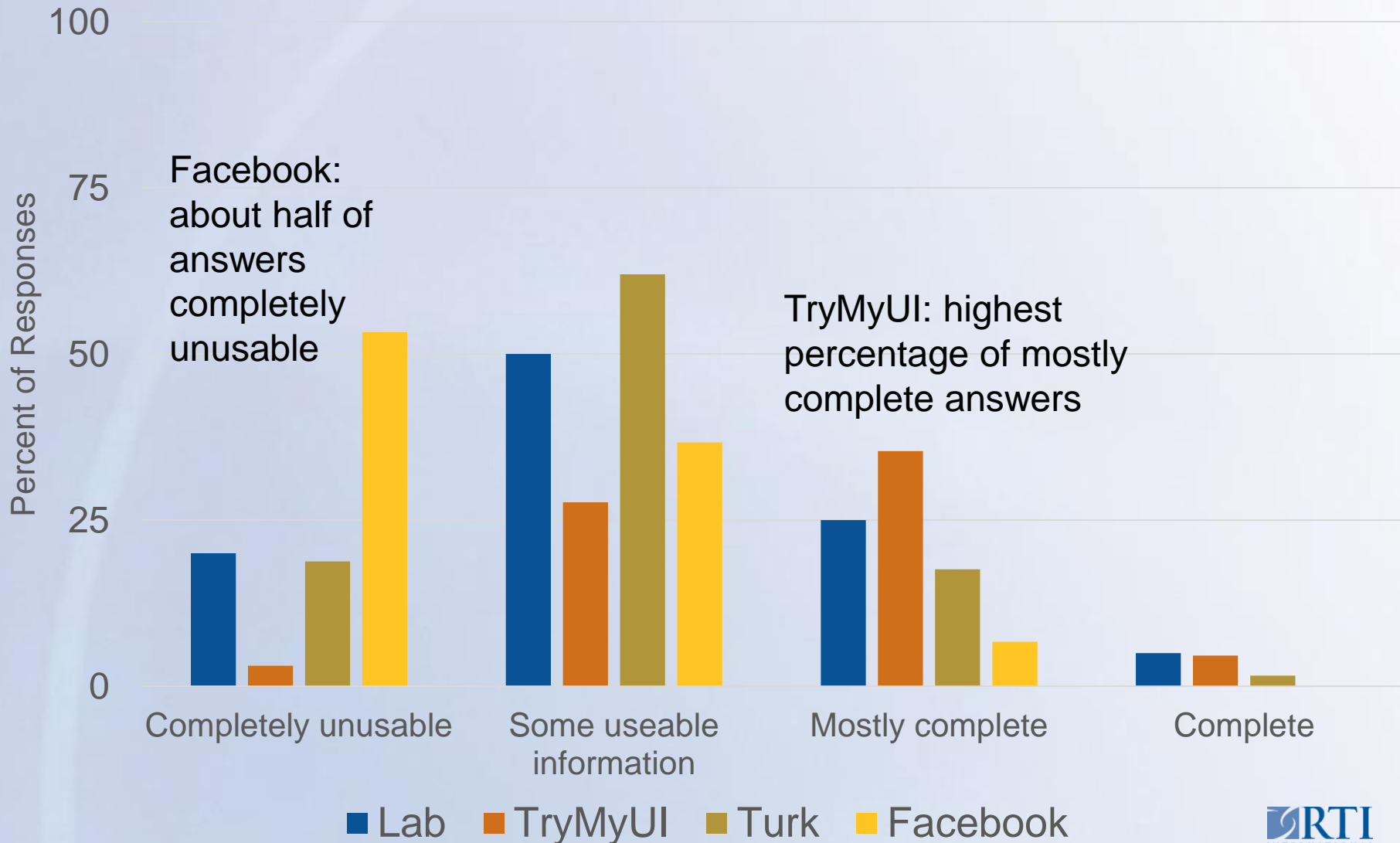
- **Somewhat usable:** “I bought two pairs of shoes and they were \$50 a pair, so I came up with a \$100.”

Results: response strategy quality, examples

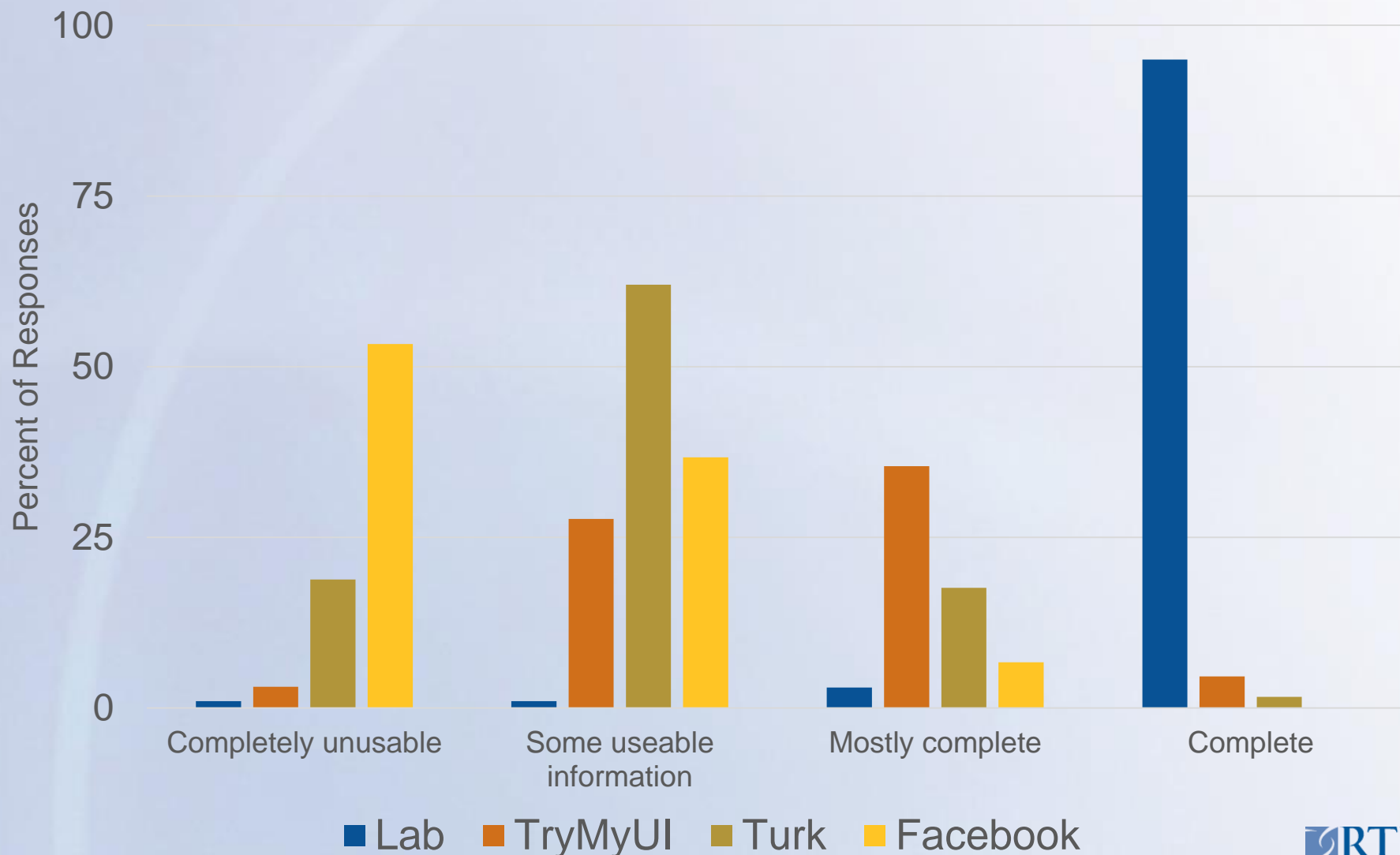
“Since the first of {reference month} how much have you or any member of your household spent on clothing?... *How did you arrive at your answer?*”

- **Complete:** “We did quite a bit of back to school shopping and I was just trying to come up with a number, cause there was quite a bit, I have two children. So I just roughly said, probably about 200 for each child is my guess online. Just I was going, website, by website. There were two main websites, well three websites, so there was LL bean and Lance End. And I just basically divided it up, and there was a little bit on Children’s Place. So, I remember spending around 80 on the Children’s Place with leaving about 320 for the rest. And I thought yeah that would be about right. You know 200 at Lance End and the other 120 at LI Bean. That seemed about right for me. I was just trying to come up with numbers.”

Results: initial response strategy quality by platform



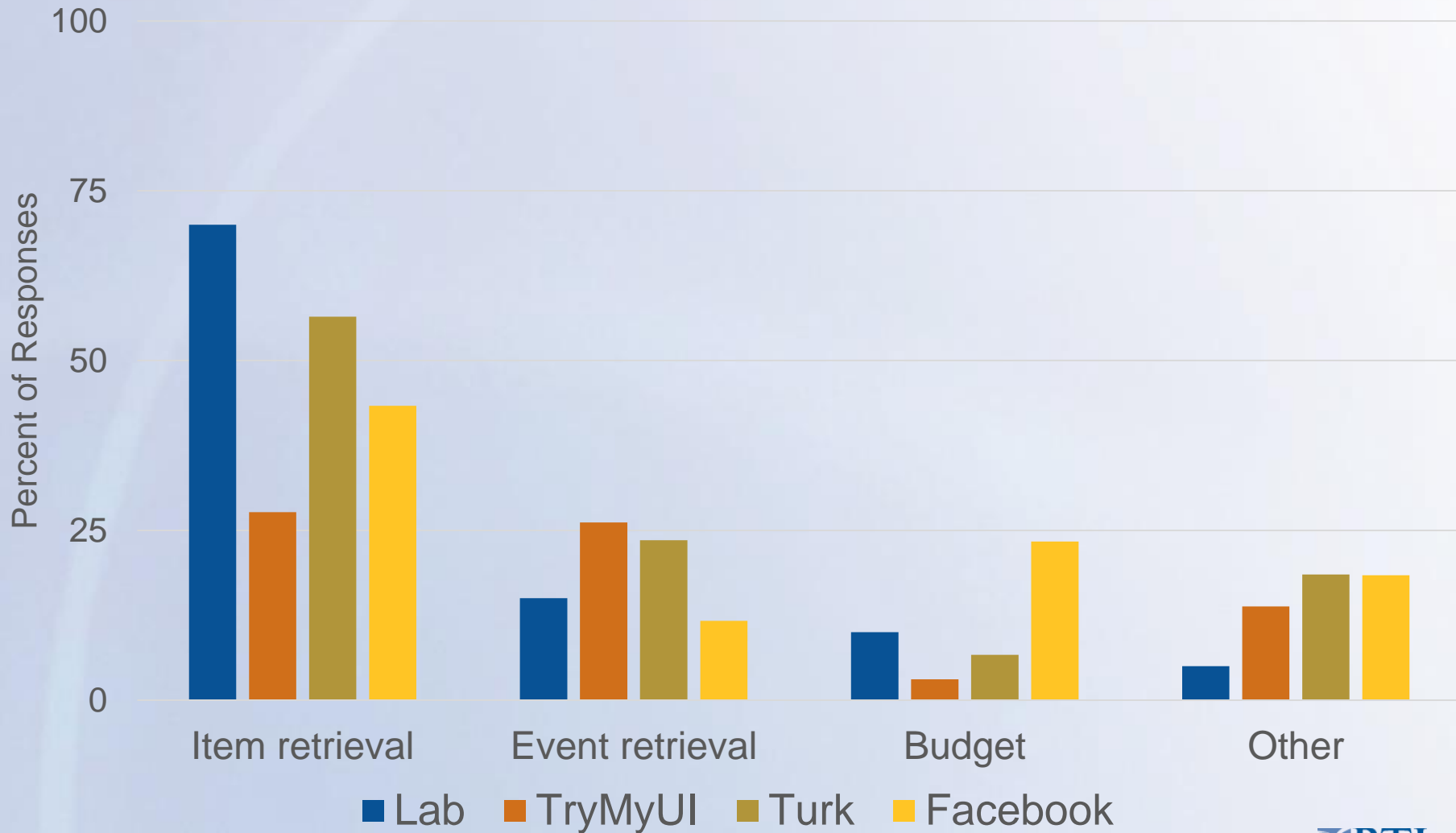
Results: response quality with follow-up probes



Results: response strategies

Item retrieval	participants retrieve information about specific items and report the sum of those events
Event retrieval	participants use information from specific events (shopping trips) and report the sum of those
Budget	participants use their planned budget number as a response, or use their budget as a basis for response
Other	retrieval and estimation, guessing, general Impression, receipts, misc.

Results: response strategies



Conclusions

- Both traditional and crowdsourcing methods allowed us to evaluate comprehension
- Differential success measuring response strategy
 - Almost half of Facebook responses did not provide useable information
 - The verbal modes (lab & TryMyUI) captured more useable information
 - Facebook and MTurk participants tended to just answer the questions asked
 - Ability to probe further important

Advantages of crowdsourcing

- Fast
- Cheap
- Geographic dispersion
- Experienced audience (esp. TryMyUI)
- Can target specific groups (e.g. Facebook)

Disadvantages of crowdsourcing

- Lack of follow up probes
 - Simple CI tasks may not need follow up (e.g. comprehension)
 - Follow-up may be more important for more complex tasks (e.g. response strategies)
- Potential panel bias
 - If panel members different from population
 - May be similar concerns with lab participants

Implications

- Crowdsourcing can be a viable “fit for use” recruiting method
 - Particularly when there may be regional differences
 - Allows for larger samples faster and cheaper than lab work

- Self-administration may be a viable CI method for simple tasks

Recommendations

- Consider using crowdsourcing as a part of your CI studies
 - During preliminary stages to gain insight into concept comprehension
 - Or final stages to larger-scale testing of proposed wording
- Incorporate with traditional lab CI to collect more in-depth information
- Carefully consider demographics & experiences, use both methods to optimize quality/quantity

Future research

- Further evaluate each mode and mixed designs for optimal utility
- Investigate additional crowdsourcing alternatives (e.g. Google Consumer Surveys, Promoted Tweets, other panels like TryMyUI)
- More work to figure out when and where you need an interviewer

Questions?

Joe Murphy jmurphy@rti.org  [@joejohnmurphy](https://twitter.com/joejohnmurphy)


Jennifer Edgar edgar.jennifer@bls.gov

Michael Keating mkeating@rti.org



SurveyPost

<http://blogs.rti.org/surveypost>

 [@SurveyPost](https://twitter.com/SurveyPost)

References

- Edgar, J. (2013). Self-Administered Cognitive Interviewing. Presented at the 68th Annual Conference of the American Association for Public Opinion Research, Boston, MA.
- Keating, M., Rhodes, B., & Richards, A. (2013). Crowdsourcing: A Flexible Method for Innovation, Data Collection, and Analysis in Social Science Research. In *Social Media, Sociality, and Survey Research*, (eds. Hill, C.A., Dean, E., & Murphy, J.). New York: Wiley.
- King, S. (2009). Using Social Media and Crowd-Sourcing for Quick and Simple Market Research. <http://money.usnews.com/money/blogs/outside-voices-small-business/2009/01/27/using-social-media-and-crowd-sourcing-for-quick-and-simple-market-research>
- Murphy, J. (2013). Altruism: Alive and Well on Facebook? *SurveyPost*. <http://bit.ly/Hjyrwl>
- Murphy, J., Sha, M., Flanigan, T. S., Dean, E. F., Morton, J. E., Snodgrass, J. A., & Ruppenkamp, J. W. (2007). Using Craigslist to recruit cognitive interview respondents. Presented at the Annual Meeting of the Midwest Association for Public Opinion Research, Chicago, IL.



Supplemental slides

Past research

- Using Crowdsourcing methods to collect “cognitive interviewing” type data (Edgar, 2012; Edgar, 2013, Murphy et al, 2013)
- Promising results in terms of
 - Participant characteristics
 - Cost efficiency
 - Time efficiency
- Preliminary results suggest data collected via crowdsourcing may be comparable to data collected in the lab
 - Relevance of responses
 - Comprehension of target concept

Methods summary

	Traditional CI	TryMyUI	Facebook	MTurk
Recruiting	Quota + convenience sample	Quota sample	Targeted English speaking; “like” music; “like” Red Cross	Open to everyone
Location	Four cities	National Panel		
Duration	20-30 minutes	Limited to 20 minutes	~10 minutes	~5 min modules
Questions	All verbal responses	Typed survey response, verbal probe response	Typed survey response and typed probe response	
Sample size	71	44	60	250 per module
Data capture	Notes and audio recording	SurveyMonkey, audio recording of talk aloud	SurveyGizmo	