

## Appendix C: Reliability analysis for open-ended codes

A senior member of the research team manually coded more than 360,000 open-ended answers using the coding protocol in Appendix B. To evaluate the reliability of these codes, a sample of study interviews was selected and independently coded by a team of five researchers. This sample consisted of all 6,940 respondents for which one or more open-ended answers was initially coded as gibberish, item nonresponse or non sequitur. An additional 500 cases whose answers were all coded as either responsive or blank were randomly selected from each of the six samples. In total, 57,599 open-ended answers from 9,940 respondents were used to measure the reliability of the coding scheme.<sup>27</sup> The sample was randomly divided into four equal sized batches, and each batch was coded by two different researchers such that each open-ended answer was independently coded a total of three times including the initial code.

A coding scheme is said to be reliable when different people following the same set of instructions tend to agree on the proper classification of the answers to be coded. If coders frequently disagree about the classification, then the coding would be unreliable. Across all questions and answers in this study, the reliability

coders reached the same conclusion as the primary coder 94% of the time. The agreement rate varied somewhat by question, ranging from 92% for GETDONE to 96% for GREWUPCITY and RETIRE.

In addition to calculating the chance of agreement with the initial coding, Krippendorff's alpha was also computed to measure the chance-adjusted

probability of agreement between all four coders.<sup>28</sup> Krippendorff's alpha was chosen as a reliability metric for this analysis because it accommodates multiple coders and allows for the possibility that not every answer will have been coded by the same set of people. For a given question, the

---

### Reliability scores for coding of open-ended questions

*Intercoder reliability scores before adjudication*

Question	Percent agreement (%)	Krippendorff's alpha
CITYVISIT	95	.81
COMPUTER	93	.84
FEELS	96	.77
GETDONE	92	.87
GREWUPCITY	96	.85
RETIRE	96	.81
<b>Overall</b>	<b>94</b>	<b>.85</b>

Notes: See Appendix B the full wording of the open-ended questions

Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019. "Assessing the Risks to Online Polls from Bogus Respondents"

PEW RESEARCH CENTER

---

<sup>27</sup> Answers that were left blank by the respondent were coded automatically. Because these did not involve any individual discretion or judgment to code, they were excluded from the reliability analysis.

<sup>28</sup> Krippendorff, K. 2004. "Reliability in Content Analysis: Some Common Misconceptions and Recommendations." *Human Communication Research* 30(3):411-33.

value of alpha ranges from 0 to 1, where 0 means that coders always disagree or are assigning codes randomly and 1 means that coders always agree on the correct classification.

While there is no one-size-fits-all threshold, an alpha of 0.8 is generally considered to be desirable. Taking all of the questions and answers together, the codes have an alpha of 0.85. Individually, all but one of the questions had an alpha of 0.8 or higher. The exception was FEELS which had an alpha of .77. After the reliability coding was completed, answers where two or more of the reliability coders disagreed with the primary coder were reviewed and a final code was chosen.