

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Sociology Department, Faculty Publications

Sociology, Department of

2022

“Are You ...”: An Examination of Incomplete Question Stems in Self-administered Surveys

Nestor Hernandez

Kristen Olson

Jolene D. Smyth

Follow this and additional works at: <https://digitalcommons.unl.edu/sociologyfacpub>



Part of the [Family, Life Course, and Society Commons](#), and the [Social Psychology and Interaction Commons](#)

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Sociology Department, Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

“Are You ...”: An Examination of Incomplete Question Stems in Self-administered Surveys

Nestor Hernandez, Kristen Olson, and Jolene D. Smyth

University of Nebraska-Lincoln, Lincoln, NE, USA

Corresponding author — Nestor Hernandez, Department of Sociology,
University of Nebraska-Lincoln, 711 Oldfather Hall, Lincoln, Nebraska, USA.
Email: nestor.hernandez@huskers.unl.edu

ORCID Kristen Olson <https://orcid.org/0000-0002-8004-0226>

Abstract

Questionnaire designers are encouraged to write questions as complete sentences. In self-administered surveys, incomplete question stems may reduce visual clutter but may also increase burden when respondents need to scan the response options to fully complete the question. We experimentally examine the effects of three categories of incomplete question stems (incomplete conversational, incomplete ordinal, and incomplete nominal questions) versus complete question stems on 53 items in a probability webmail survey. We examine item nonresponse, response time, selection of the first and last response options, and response distributions. We find that incomplete question stems take slightly longer to answer and slightly reduce the selection of the last response option but have no effect on item nonresponse rates or selection of the first response option. We conclude that questionnaire designers should follow current best practices to write complete questions, but deviations from complete questions will likely have limited effects.

Published in *Field Methods* 2022

DOI: 10.1177/1525822X221134756

Copyright © 2022 Nestor Hernandez, Kristen Olson, and Jolene D. Smyth. Published by SAGE Publications. Used by permission.

Introduction

Survey designers generally agree that a well-written survey question is clear about the question's focus and what kind of answer is desired and acceptable (Bradburn et al. 2004; Dillman et al. 2014; Saris and Gallhofer 2007; Schaeffer and Dykema 2020), yet the role that different parts of a question play to communicate a question's focus and desired answer has received surprisingly little empirical attention. Questionnaire designers acknowledge that questions have multiple parts, with the two most important being the statement of the question itself [(Blair et al. 2013); also called the "request for answers" (Saris and Gallhofer 2007) or the "question stem" (Dillman et al. 2014)] and the response options. Questionnaire design best practice recommendations are often unclear about how complete the question statement itself needs to be. For instance, Dillman et al. (2014) advise: "Use complete sentences that take a question form, and use simple sentence structures" (p. 121). Blair et al. (2013) say: "A closed question has two parts: the statement of the question and the response categories. We need to devote equal attention to each part." (p. 193). Schaeffer and Dykema (2020) state: "A basic principle is that questions should project the format of the response" (p. 52).

The distinction between question stems and response options varies by mode. In an interviewer-administered survey, the distinction between the question and the response options is visible only to the interviewer (Bradburn et al. 2004). In fact, interviewer-administered modes often have question stems that are separated from the response options; the questions then are completed when the interviewer follows the instructions to read the response options. In self-administered surveys, respondents control how the question is read. In a self-administered survey in particular, the distinction between the question and the response options is visible to the respondent, thus the issue of what the question stem is supposed to accomplish becomes important. To save space and to follow recommendations to use as few words as needed in self-administered surveys, researchers often shorten question stems, using an incomplete sentence with ellipses or a colon at the end.

These typographical conventions imply that the response options communicate the question's focus and should be used to complete the sentence (Converse and Presser 1986; Dillman et al. 2014).

We define *incomplete question stems* as questions that replace the response options in the question stem with ellipses (...) or a colon (:) and/or require the response options to be read to otherwise form a complete sentence. Incomplete question stems may be an example of what Smyth and Olson (2018) refer to as mismatches in which the "question wording does not effectively communicate what type of answer constitutes an adequate answer given the provided response options" (p. 35) compared to a matched question stem that adequately reflects all of the presented response options.

In self-administered surveys, whether a completed sentence in the question stem affects the quality of answers has received surprisingly little empirical attention. Dillman et al. (2014) provide an example of a two-word prompt "your county" that may be misread as "country." They argue that writing the prompt as a complete question would allow the researcher to provide more information that would help respondents understand what was being asked of them, suggesting, "In what Idaho county do you live?" Smyth and Olson (2018) found that there was no difference overall in response distributions for an incomplete question stem ("Is your home...") experimentally compared to a complete question stem that emphasized only one of the response options, but that there was higher endorsement of the emphasized response option in the complete (but mismatched) question stem.

In this article, we experimentally examine the following research question: *Do incomplete question stems affect the quality of answers to survey questions?* We examine multiple indicators of response quality across three different types of incomplete question stems in a self-administered web and mail general population survey.

Types of Incomplete Question Stems

The first type of incomplete question stem does not change the main part of the question but lengthens the question stem by

including a conversational “Would you say...” to transition into the response options (see **Figure 1**). In *incomplete conversational question stems*, the domain of interest and area of interrogation (Tourangeau et al. 2000) is listed in a complete sentence in the main part of the question stem (e.g., how often, how satisfied, how much), possibly indicating multiple outcomes that will be listed in the response options. For instance, in the 2021 California Health Interview Survey, web respondents are asked “How often were you able to get an appointment within 2 days? Would you say ...” with visually separated response options of “never, sometimes, usually, and always” (UCLA Center for Health Policy Research 2021). In telephone surveys, “would you say” functions as a conversational cue to help interviewers administer the question (Dillman et al. 2014), transitioning between the main question and the response options. In particular, “would you say” tells respondents that an interviewer is about to read the response options (Blair et al. 2013; Dillman et al. 2014). This formulation is sometimes used in self-administered surveys—as in this example from the California Health

Complete	Incomplete
Type 1: Incomplete Conversational Question Stems	
<p>How busy are you on a typical day?</p> <ul style="list-style-type: none"> <input type="radio"/> Extremely busy <input type="radio"/> Very busy <input type="radio"/> Somewhat busy <input type="radio"/> A little busy <input type="radio"/> Not busy at all 	<p>How busy are you on a typical day? Would you say...</p> <ul style="list-style-type: none"> <input type="radio"/> Extremely busy <input type="radio"/> Very busy <input type="radio"/> Somewhat busy <input type="radio"/> A little busy <input type="radio"/> Not busy at all
Type 2: Incomplete Ordinal Question Stems	
<p>Do you think race relations in the <u>United States</u> are generally good or generally bad?</p> <ul style="list-style-type: none"> <input type="radio"/> Generally good <input type="radio"/> Generally bad 	<p>Do you think race relations in the <u>United States</u> are...</p> <ul style="list-style-type: none"> <input type="radio"/> Generally good <input type="radio"/> Generally bad
Type 3: Incomplete Nominal Question Stems	
<p>Which of the following best describes your relationship status?</p> <ul style="list-style-type: none"> <input type="radio"/> Married <input type="radio"/> Unmarried, living with a partner <input type="radio"/> Widowed <input type="radio"/> Divorced <input type="radio"/> Separated <input type="radio"/> Single, never married 	<p>Are you...</p> <ul style="list-style-type: none"> <input type="radio"/> Married <input type="radio"/> Unmarried, living with a partner <input type="radio"/> Widowed <input type="radio"/> Divorced <input type="radio"/> Separated <input type="radio"/> Single, never married

Figure 1. Examples of incomplete question stems, Nebraska 2020.

Interview Survey—especially when borrowing question wording from interviewer-administered surveys for self-administered or mixed-mode surveys with the goal of a unified-mode design (Dillman et al. 2014).

The second type of incomplete question stem is what we call an *incomplete ordinal question stem*. In this formulation, ordinal questions are asked such that only the topic of the ordinal scale is revealed in the question stem (“Do you think the number of immigrants in America nowadays should be ...” from the web-based 2021 General Social Survey (Davern et al. 2021) and the response options reveal the concepts and quantifiers that are to be used (fully labeled from “increased a lot” to “reduced a lot”). In particular, the type of rating scale, its direction (unipolar versus bipolar) and number of scale points are only revealed through reading the response options. Because the response options are replaced by ellipses (...), this type of incomplete question stem shortens the question. Writing these questions as full questions often simply lists the complete set of response categories in the question wording itself, which some survey designers may worry creates unnecessary repetition.

The third question type is similar to incomplete ordinal questions in that it reduces the question stem length by replacing the response options with (...) but differs in that the response categories are nominal. As such, the reduced form of an *incomplete nominal question stem* is often even shorter and more ambiguous without the response categories than that of an ordinal question stem. At the extreme, these question stems are of the form “Are you ...” (or “Are you:”) where only the response categories indicate what kind of domain is being asked (gender; marital status; sexual orientation). For instance, the paper 2019 National Household Education survey asked “Is this house or apartment ...” with the response options indicating that the desired answers have to do with ownership (e.g., “owned or being bought by someone in this household”) (National Center for Education Statistics 2019). Rewriting these questions as full questions often requires a complete reformulation of the question stem.

How Incomplete Question Stems May Affect Cognitive Processing of Questions

As described by Tourangeau et al. (2000), one of the key goals in comprehending survey questions is to identify the “uncertainty space” about which the question is asking for information. Difficulties arise when the question’s focus is ambiguous or poorly specified. When question stems are incomplete, the interrogative form of a question is not fully specified. Thus, the respondent must continue processing information beyond the main question stem to fully grasp the question’s focus and domain of interrogation. *This extra complexity in the comprehension task may be burdensome and thus increase item nonresponse rates (H1) and response time (H2) for incomplete question stems compared to complete question stems.*

For respondents who process the question and proceed to give a substantive response, incomplete question stems may also change survey answers. In interviewer-administered surveys, control over reading the question is guided by an interviewer. Even when respondents interrupt, interviewers can remind respondents that the full set of response options must be read. In contrast, respondents have the locus of control in self-administered modes. In these modes, when shortcutting responses, respondents spend more time processing response options that are presented in the first half of a vertical list (Galesic et al. 2008) and the left side of a horizontal list (Menold et al. 2014), although this visual processing of early items in a list applies primarily to attitudinal questions rather than factual questions (Neuert 2017). We know of no published studies that use eye tracking to evaluate incomplete question stems but anticipate that incomplete question stems will encourage cognitive shortcuts to visually process response options. If incomplete question stems encourage incomplete processing for attitudinal questions, *we anticipate higher rates of selecting answers in the top part of a vertical set of response options and on the left side of a horizontal set of response options (H3) when incomplete question stems are presented (and thus lower rates of answers at the end of the list, H4) compared to when complete question stems are presented. We do not expect any differences for factual questions.*

The effect of an incomplete question stem may vary across incomplete question stem types. First, for *incomplete conversational question stems*, adding the transitional words “Would you say” to an already complete sentence provides no additional information about a question’s domain. Thus, compared to the same question that omits this conversational phrase, we expect no differences in item nonresponse (H1a). Due to the slightly longer question stem and additional visual clutter, we expect response times to increase slightly (H2). The addition of “would you say” may turn each response option into an implicit “yes/no” question (“would you say... never?”). If this occurs, response options presented early in the list may be considered more completely than options presented later in the list, and thus we would expect higher rates of selection of earlier response options (H3) and lower rates of selection of later response options (H4) and differences in response distributions.

For *incomplete ordinal question stems*, the separation of the question stem from the rating scale used in the response options may increase the perceived complexity of the question. As such, item nonresponse rates may be higher for incomplete ordinal questions compared to a question that contains the full set of response options (H1). On the other hand, the shorter question stem does not contain redundant information and reduces the visual clutter on a screen, possibly reducing the perception of burden or unnecessary information, and thus possibly reducing item nonresponse rates compared to the complete question stem version of the same question (H1b). These two processes may also cancel each other out, resulting in no clear difference in item nonresponse rates.

If not viewed as more burdensome, we expect response times to be shorter for incomplete ordinal question stems than for complete question stems because words are omitted (H2b). We expect differences in the distribution of responses between questions with incomplete versus complete ordinal question stems. With complete questions, all of the response options are presented before the request for an answer and thus are likely to be fully processed before one begins marking an answer. However, with incomplete question stems, the response options will be processed at the same time that they are read and that the request for an answer is made. In the latter case, we expect that respondents will be more likely to

select the first response option that is reasonable for them rather than the best answer. Therefore, we expect higher rates of endorsement of response categories that appear in the top part of a vertical scale or on the left side of a horizontal scale in questions with incomplete ordinal question stems because respondents are selecting the first option that applies to them (H3). We expect lower rates of endorsement of the last response option because respondents are likely to find a suitable response option prior to getting to this option (H4).

Finally, for *incomplete nominal question stems*, the domain of the question will dictate the response effects. In this article, we examine easily retrievable autobiographical and/or demographic topics. Thus, the respondent must search for the categories that apply to them, but not engage in a more complex attitude construction task. For an easily retrievable autobiographical question that is fully defined by the response options, we expect no differences in item nonresponse rates (H1c), a shorter response time for incomplete question stems because the questions are shorter and require less reading before the search task begins (H2c), and no substantial differences in the distribution of responses (H3) or selection of different response options (H4) due to the factual nature of the questions.

To our knowledge, no study has experimentally assessed the effects of incomplete question stems on data quality and response distributions across multiple items. We examine 53 experimental comparisons from a general population survey on experiences during the early COVID-19 pandemic.

Data and Methods

The data for this study are from the Nebraska 2020 Survey (NE2020) conducted by the Bureau of Sociological Research at the University of Nebraska-Lincoln. NE2020 is a mixed-mode self-administered web and mail survey collected between September 14, 2020, and December 14, 2020. A simple random sample of 10,000 households in Nebraska was selected by Dynata from the list of addresses maintained by the U.S. Postal Service, and the adult in

the household with the next birthday was asked to complete the survey. All contact attempts were sent via postal mail; web respondents completed the survey by typing the study's URL and their unique login ID into a web browser. The total respondent sample size is 2,811, with 1,157 respondents completing the web version and 1,654 respondents completing (AAPOR 2016).

Addresses were randomly assigned to one of two experimental versions of the 12-page questionnaire on Nebraskans' attitudes toward the COVID-19 pandemic and recent racial unrest. The incomplete question stem experiment was administered on 23 single item questions and four batteries for a total of 53 separate items across the questionnaire in a within-subjects design; respondents who received version A of the questionnaire received some questions with complete stems; others with incomplete stems and respondents who received version B had the complement set of questions with complete and incomplete question stems. There was no difference in characteristics of respondents across the two questionnaire versions ($p > 0.05$, Online Appendix Table A.1). For these 53 items, only the question stems changed; response options were identical across the incomplete and complete versions. We used a unified mode design to minimize any potential differences in presentation of items across modes.

Three types of incomplete question stems are examined (Figure 1). The *incomplete conversational question stem* extends the main question by including "Would you say ...?" In our experiment, incomplete conversational question stems include two single item questions and four battery items (34 separate items), for a total of 36 separate items.

Incomplete ordinal question stems design replaces the ordinal response options that are included in the question stem with ellipses. Eleven items are included as ordinal incomplete question stems.

Incomplete nominal question stems replace a substantial portion of the question stem with ellipses, changing the question stem to facilitate the absence of nominal response options. In these questions, the respondent may not know what construct is being queried from the question stem alone. Six items are included as nominal incomplete question stems.

Dependent Variables

Item nonresponse is a dichotomous variable, where 1 indicates that a respondent did not answer the question and 0 indicates that a substantive answer was provided. To avoid conflating breakoffs with item nonresponse, we examine item nonresponse rates among respondents who had at least provided an answer to the immediately preceding question (e.g., item nonresponse on Q45 examined on respondents who had not broken off on or before Q44). Six respondents broke off before answering questions included in this experiment and have been excluded from the analysis ($n = 2,805$). The average item nonresponse rate across questions is 3.1%.

Response time is calculated for respondents who completed the web questionnaire (response times are unavailable for mail respondents). Response times are measured as the number of seconds spent answering a survey question, operationalized as (the time the respondent submitted their answer) minus (the time the respondent first entered the page). Time spent on the battery items is measured at a screen level, not at a subitem level. To account for nonnormally distributed response times and outliers, response times were trimmed to the 1st and 99th percentile and log-transformed (Yan and Olson 2013). The average logged response time across questions is 0.97 log-seconds.

As a measure of *effects on survey responses*, we examine the proportion of respondents who select the first or the last response option for each item. Each item has a different number of response options, ranging from two to five. Thus, only selecting the first option can be defined equivalently across questions. Selecting the last option also can be defined equivalently across items. The average proportion of first response options selected across questions is 23.1%, and the average proportion of last response options selected is 15.6%.

We also examine *response distributions* for each item across the experimental versions. To do this, we examine the full distribution in each answer category.

Independent Variable

The key independent variable is the question-level experimental treatment of complete versus incomplete question stem.

Analysis

We have 53 repeated measures on each respondent across the three categories of incomplete question stems. To account for repeated measures, we estimate cross-classified random effects models for our data quality indicators of item nonresponse ($n = 142,981$ respondent questions), response time ($n = 24,882$ respondent questions), and selection of the first and last response options ($n = 138,158$ respondent questions). We are interested in the effect of incomplete question stems overall (*Incomplete*) and for each type of incomplete question stem (*QnType* = Conversational (reference), Ordinal, or Nominal), and thus include an interaction effect between the type of incomplete question stem and the experimental variation indicator. We include a grand-mean centered nonresponse-adjusted sampling weight (*SurveyWeight*) (Snijders and Bosker 2012) and the mode of data collection (*Mode* = mail vs. web) as control variables. Thus, for respondent i and question j , where $f(Y_{ij})$ indicates the appropriate link function (logit for item nonresponse, select first, and select last; linear for response time), $u_i \sim N(0, \tau_R^2)$, $u_j \sim N(0, \tau_Q^2)$, and $e_{ij} \sim N(0, \sigma^2)$ (for the linear model), we estimate

$$\begin{aligned}
 f(Y_{ij}) = & \beta_0 + \beta_1 (Incomplete = 1)_{ij} \\
 & + \beta_{2a} (QnType = Ordinal)_j \\
 & + \beta_{2b} (QnType = Nominal)_j \\
 & + \beta_{3a} (QnType = Ordinal)_j (Incomplete = 1)_{ij} \\
 & + \beta_{3b} (QnType = Nominal)_j (Incomplete = 1)_{ij} \\
 & + \beta_4 SurveyWeight_i \\
 & + \beta_4 (Mode = Mail)_i \\
 & + u_i + u_j + e_{ij}
 \end{aligned}$$

We test the joint hypothesis that β_{3a} and β_{3b} equal zero to test whether the effect of incomplete question stems varies across the types of questions. To ease interpretation, we present results from the full models as predicted margins and differences between incomplete and complete question stems as average marginal effects (Mize 2019). Full model results are presented in the supplementary materials (Online Appendix Tables A.2 and A.3).

We also examine response distributions across each of the 53 items, using survey design-adjusted F -tests with Taylor Series Linearization using Stata 17. We use a Bonferroni correction ($p < 0.05/53 = 0.0009$) to indicate statistical significance for these item-level analyses (see Online Appendix A.4 for full results).

Findings

Table 1 presents predicted values and average marginal effects from the multilevel analyses.

Item Nonresponse Response Time

We examined response times for web respondents. Consistent with the burden hypothesis (H2), when all of the question types are combined, there is a small statistically significant increase in response time for questions with incomplete question stems in the web version of the survey (Complete: 1.01 log-seconds; Incomplete: 1.02 log-seconds, $\exp(0.013) = 1.01$ second difference; $p = 0.045$). We find no significant differences in response time across the types of incomplete question stems ($\chi^2 = 1.47$, $p = 0.48$), inconsistent with H2b and H2c.

Selecting the First and Last Response Options

Next, we examined whether respondents tended to move toward selecting the first or away from selecting the last response option. Inconsistent with H3, there was not a shift toward the first

response option overall (Complete: 19.6%; Incomplete: 19.6%; $p = 0.99$). However, there was a statistically significant shift away from the last response option with incomplete question stems (Complete: 8.3%, Incomplete: 7.9%; $p = 0.01$), consistent with H4. There was no difference across the types of incomplete question stems for selection of the first ($\chi^2 = 1.12$, $p = 0.57$) or last ($\chi^2 = 0.82$, $p = 0.66$) response option.

Response Distributions

Finally, we examine whether the distribution of responses differs for questions with incomplete question stems compared to complete question stems. In aggregate, the multilevel models indicate that there are shifts away from the last response option. For individual items, only seven (13%) out of 53 total items differed in their distribution of responses at $p < .05$, and only three at the $p < .0009$ level, generally following the expected pattern for these three attitudinal items. Among the 36 incomplete conversational question stem items, 33 (92%) showed no differences at the $p < .05$ level between the complete and incomplete question stems, and only one item met the $p < .0009$ criteria. On the item about adequacy of community resources, respondents in the incomplete conversational question stem condition were more likely to say that fire resources were “too little” in their community (Incomplete: 22.6%, Complete: 17.3%) and less likely to say that resources were “too much” (Incomplete: 2.8%, Complete: 5.2%). For the 11 incomplete ordinal stems, there are statistically significant differences between complete and incomplete question stems at the $p < .0009$ level for two (18%) items. For Q47 ($F = 6.82$, $p < 0.0001$), respondents were more likely to say that race relations in Nebraska are “generally bad” when the complete question stem is provided (Incomplete: 27.2%, Complete: 33.6%). The overall response distribution also differed in responses to how serious of a problem police brutality is in the United States (Q49, $F = 5.66$, $p < 0.0001$). There was higher selection of the first response option “a very serious problem” for incomplete question stems (Incomplete: 37.0%; Complete: 29.5%), but also higher selection of the last response

option “not a problem at all” (Incomplete: 9.2%; Complete: 6.2%). Finally, for the six incomplete nominal question stem items, as expected, none of the response distributions varied across experimental conditions ($p > 0.0009$).

Discussion

Despite their widespread use, the effects of incomplete question stems on the quality of survey responses have received relatively little empirical attention. We examined whether incomplete question stems in self-administered surveys alter response behaviors, perhaps by assisting respondents by removing visual clutter or by adding burden for respondents when they track between the question and the response options to complete a question.

In general, there were few differences in our data quality outcomes across 53 questions written with complete versus incomplete question stems. This is largely good news for survey designers. There were no differences in item nonresponse rates or selection of the first response option. There were, however, differences in response times among web survey respondents— incomplete question stems required about a second longer for web survey respondents to answer. This pattern of results is consistent with our hypothesis of increased burden for incomplete question stems. Additionally, respondents tended to select the last response option less often, suggesting that the responses were moved toward the earlier response categories, consistent with our hypothesis that incomplete question stems may shortcut processing of later response options.

We divided incomplete question stems into three types and found similar patterns of effects. Interestingly, all three items that showed differences in response distributions across experimental treatments were attitudinal items, not incomplete nominal questions. However, these differences were quite modest and would not change substantive conclusions. Further research should explore the types of incomplete items and attitude domains that are most affected by this design decision.

As with any study, this study has limitations. First, many of our incomplete conversational question stems were displayed in

grids; we did not see notable differences across the single items and grids, but future work should examine the effects of incomplete conversational question stems on other single-item questions. Second, we only examined six incomplete nominal questions; these items reflect a range of autobiographical questions often asked with incomplete question stems, but future research should examine other questions. Second, our study was conducted on a statewide probability sample of Nebraskans in a survey about COVID-19 and race relations. We have no reason to expect that the mechanisms for complete or incomplete question stems would be different in a national sample or would interact with the question topic (beyond expected differences for attitudinal versus factual questions), but future research should evaluate this directly. Third, we did not examine whether the effects of incomplete question stems varied across mail and web respondents. Disentangling differences in the composition of the respondents in each mode to assess measurement differences is beyond the scope of this paper. Additionally, this survey was administered only in English, omitting about 7% of Nebraska residents who do not speak English at least “very well” (U.S. Census Bureau 2020). Future research should examine the effects of incomplete questions in multilingual surveys. Finally, we see evidence of shifts in response distributions away from the last response option, but do not know which version provides the more accurate answer. This is true with any study of attitudinal items; future work could evaluate questions for which “true value” validation data exist, but these are the factual items for which we would expect the fewest effects of incomplete question stems.

Questionnaire designers are advised to write questions as complete sentences, and we find no evidence to change that advice. Incomplete question stems yield slightly longer response times and slightly reduce the selection rate of the last response option. However, the magnitude of the effects is small, on average. Thus, questionnaire designers looking to save space in self-administered surveys with incomplete questions or who retain a conversational “would you say” to align across interviewer and self-administered modes should expect only limited effects on answers to survey questions.

* * * * *

Conflicting interests The authors have no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Funding Data collection for the NE2020 Survey was supported by internal research funds.

Supplemental Material follows the **References**.

References

- American Association for Public Opinion Research (AAPOR). 2016. Standard definitions final dispositions of case codes and outcome rates for surveys. [https://www.aapor.org/Standards-Ethics/Standard-Definitions-\(1\).aspx](https://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx) (accessed September 30, 2022).
- Blair, J., R. F. Czaja, and E. A. Blair. 2013. *Designing surveys: A guide to decisions and procedures*, 3rd ed. Los Angeles: Sage.
- Bradburn, N. M., S. Sudman, and B. Wansink. 2004. *Asking questions: The definitive guide to questionnaire design-for market research, political polls, and social and health questionnaires*. San Francisco: Jossey-Bass.
- Converse, J. M., and S. Presser. 1986. *Survey questions: Handcrafting the standardized questionnaire*. Thousand Oaks, CA: Sage.
- Davern, M., R. Bautista, J. Freese, S. L. Morgan, and T. W. Smith. 2021. *General social survey 2021 cross-section*. [machine-readable data file] edited by NORC. Chicago.
- Dillman, D. A., J. D. Smyth, and L. M. Christian. 2014. *Internet, phone, mail, and mixed mode surveys: The tailored design method*. Hoboken, NJ: John Wiley & Sons.
- Galesic, M., R. Tourangeau, M. P. Couper, and F. G. Conrad. 2008. Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly* 72:892–913.
- Menold, N., L. Kaczmarek, T. Lenzner, and A. Neusar. 2014. How do respondents attend to verbal labels in rating scales? *Field Methods* 26:21–39.
- Mize, T. D. 2019. Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science* 6:81–117.
- National Center for Education Statistics. 2019. *National household education survey: Early childhood program participation questionnaire*, edited by National Center for Education Statistics. Washington, DC.

- Neuert, C. E. 2017. Processing forced-choice versus check-all-that-apply question formats: Evidence from eye tracking. *Field Methods* 29:383–94.
- Saris, W. E., and I. N. Gallhofer. 2007. *Design, evaluation, and analysis of questionnaires for survey research*. Hoboken, NJ: John Wiley and Sons.
- Schaeffer, N. C., and J. Dykema. 2020. Advances in the science of asking questions. *Annual Review of Sociology* 46:37–60.
- Smyth, J. D., and K. Olson. 2018. The effects of mismatches between survey question stems and response options on data quality and responses. *Journal of Survey Statistics and Methodology* 7:34–65.
- Snijders, T. A. B., and R. J. Bosker. 2012. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Los Angeles: Sage.
- Tourangeau, R., L. J. Rips, and K. A. Rasinski. 2000. *The psychology of survey response*. Cambridge: Cambridge University Press.
- UCLA Center for Health Policy Research. 2021. California health interview survey 2021: Adult CAWI questionnaire, edited by UCLA Center for Health Policy Research, California Department of Health Care Services and California Department of Public Health. Los Angeles.
- U.S. Census Bureau. 2020. *S1601: Language spoken at home, 2019 American community survey 1-year estimates*. Suitland, MD: U.S. Census Bureau.
- Yan, T., and K. Olson. 2013. Analyzing paradata to investigate measurement error. In *Improving surveys with paradata: Analytic uses of process information*, ed. F. Kreuter, 73–96. Hoboken, NJ: John Wiley & Sons.

Supplementary Materials:
“Are you...”: An Examination of Incomplete Question Stems in Self-Administered Surveys
Nestor Hernandez, Kristen Olson, and Jolene D. Smyth
University of Nebraska-Lincoln

Table A.1: Demographic Distribution of Respondents across Experimental Versions

	Version A	Version B	Overall	<i>p</i> -value
Education				
High School or less	33.14%	32.63%	32.88%	0.73
Some College	36.86%	35.56%	36.20%	
College grad or more	30.01%	31.80%	30.92%	
Age				
19–44	46.16%	44.31%	45.22%	0.43
45–64	32.11%	35.05%	33.60%	
65+	21.74%	20.64%	21.18%	
Gender				
Male	48.29%	49.63%	48.97%	0.62
Female	51.71%	50.37%	51.03%	
Race				
White	85.42%	88.76%	87.13%	0.12
Other races	14.58%	11.24%	12.87%	
Mode				
Web	44.50%	40.25%	42.38%	0.09
Mail	55.50%	59.75%	57.62%	

Note: Sample size ranged from 2,594 to 2,811 across variables due to missing values. All analyses account for complex sample design.

Table A.2 Coefficients and Standard Errors from Multilevel Logistic and Linear Models Predicting Item Nonresponse and Response Time, NE2020

	Logistic: Item Nonresponse				Linear: Response Time			
	Model 1: Main Effects		Model 2: Interaction Effects		Model 1: Main Effects		Model 2: Interaction Effects	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Incomplete Question Stem = 1	0.083*	0.042	0.073	0.050	0.013*	0.007	0.001	0.013
Incomplete Question Type								
Ordinal	-0.123	0.277	-0.109	0.281	-1.972****	0.355	-1.980****	0.356
Nominal	-0.671	0.357	-0.739*	0.365	-1.859****	0.404	-1.869****	0.404
Incomplete Stem * Incomplete Type								
Incomplete * Ordinal			-0.027	0.088	--	--	0.015	0.016
Incomplete * Nominal			0.131	0.144	--	--	0.021	0.019
Survey weight	-0.0004**	0.0001	-0.0004**	0.0001	-0.0001****	0.00002	-0.0001****	0.0000
Mode Mail=1	1.729****	0.138	1.729****	0.138	--	--	--	--
Intercept	-7.207****	0.200	-7.207****	0.200	2.421****	0.286	2.427****	0.286
Variance Components								
Questions	0.620	0.114	0.620	0.114	0.490	0.145	0.490	0.145
Respondents	6.324	0.386	6.323	0.386	0.105	0.005	0.105	0.005
Residual			--		0.276	0.003	0.276	0.003
Likelihood ratio test	13887.55****		13886.43****		26470.3****		26470.8****	
AIC	27324.17		27327.14		41287.71		41290.23	
Log-likelihood	-13654.1		-13653.6		-20635.9		-20635.1	
N observations	142,981		142,981		24,882		24,882	
n respondents	2805		2805		1152		1152	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$.

Table A.3 Coefficients and Standard Errors from Multilevel Logistic Models Predicting Selection of the First and the Last Response option, NE2020

	Select first				Select last			
	Model 1: Main Effects		Model 2: Interaction Effects		Model 1: Main Effects		Model 2: Interaction Effects	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Incomplete Question Stem=1	0.002	0.016	-0.008	0.021	-0.053**	0.019	-0.063**	0.022
Incomplete Question Type								
Ordinal	1.332**	0.429	1.328**	0.429	0.461	0.492	0.449	0.492
Nominal	2.777****	0.549	2.751****	0.549	-2.186**	0.640	-2.220**	0.642
Incomplete Stem * Incomplete Type								
Incomplete * Ordinal	--		0.008	0.035	--		0.025	0.040
Incomplete * Nominal	--		0.053	0.050	--		0.068	0.096
Survey weight	0.00002	0.000	0.00002	0.000	0.000****	0.000	0.000****	0.000
Mode Mail=1	-0.014	0.025	-0.014	0.025	-0.086*	0.037	-0.086	0.037
Intercept	-2.246****	0.211	-2.242****	0.211	-2.358****	0.242	-2.354****	0.242
Variance Components								
Questions	1.535	0.303	1.535	0.303	2.018	0.406	2.018	0.406
Respondents	0.230	0.011	0.230	0.011	0.669	0.026	0.669	0.026
Likelihood ratio test	24640.35****		24640.59****		21426.52****		21426.12****	
AIC	107440.7		107443.6		90267.62		90270.8	
Log-likelihood	-53712.4		-53711.81		-45125.8		-45125.4	
N observations	138,158		138,158		138,158		138,158	
n respondents	2805		2805		2805		2805	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$.

Table A.4. Response Distributions

	Incomplete	Complete	Total	Design-based F	<i>p</i>
Question Type 1: Incomplete Conversational Question Stems					
Q7. Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs...					
Most of the time	49.33	50.07	49.70	0.44	0.72
Some of the time	30.87	31.44	31.16		
Only now and then	15.20	13.19	14.19		
Hardly at all	4.60	5.30	4.95		
Q19. How busy are you on a typical day? Would you say...					
Extremely Busy	13.26	14.80	14.03	0.84	0.50
Very Busy	42.06	37.65	39.86		
Somewhat Busy	30.19	32.80	31.49		
A little Busy	9.59	9.67	9.63		
Not busy at all	4.90	5.09	4.99		
Q8a. How much do you trust each of the following to provide information about the coronavirus outbreak? Would you say...					
President Donald Trump					
Completely	13.66	13.96	13.80	1.79	0.13
Mostly	23.12	19.98	21.56		
Somewhat	16.90	13.59	15.25		
A little	10.69	11.68	11.18		
None	35.64	40.80	38.21		
Q8b. Leaders of federal public health agencies such as the Centers for Disease Control and Prevention (CDC)					
Completely	14.13	13.82	13.97	0.99	0.41
Mostly	42.33	41.78	42.06		
Somewhat	25.52	27.34	26.43		
A little	11.47	12.66	12.07		
None	6.55	4.40	5.48		
Q8c. Governor Pete Ricketts					
Completely	8.14	9.43	8.78	2.70	0.03
Mostly	28.61	24.81	26.72		
Somewhat	24.99	25.37	25.17		
A little	15.12	20.61	17.85		
None	23.14	19.79	21.48		
Q8d. Leaders of the Nebraska Department of Health and Human Services (NE DHHS)					
Completely	13.72	12.27	13.00	0.95	0.43
Mostly	40.30	42.97	41.63		
Somewhat	29.13	30.38	29.75		
A little	12.41	9.62	11.02		
None	4.44	4.76	4.60		
Q8e. Leaders of your local health department					
Completely	15.86	12.97	14.42	1.20	0.31
Mostly	40.80	45.07	42.93		
Somewhat	25.40	26.27	25.84		

	Incomplete	Complete	Total	Design-based F	<i>p</i>
A little	11.96	10.84	11.40		
None	5.98	4.84	5.41		
Q8f. Leaders of your city or town					
Completely	6.98	8.45	7.71	1.50	0.20
Mostly	33.66	35.05	34.35		
Somewhat	29.10	31.81	30.46		
A little	18.91	15.58	17.25		
None	11.35	9.11	10.24		
Q8g. Your personal doctor or another healthcare provider					
Completely	34.99	37.98	36.48	1.33	0.26
Mostly	44.22	46.09	45.15		
Somewhat	14.74	10.81	12.78		
A little	3.63	3.57	3.60		
None	2.43	1.54	1.99		
Q20a. During the past 30 days, how often did you experience each of the following? Would you say...					
You were able to take time for yourself					
Always	16.93	16.53	16.73	1.87	0.12
Often	28.05	24.13	26.09		
Sometimes	36.43	38.77	37.60		
Rarely	16.42	16.26	16.34		
Never	2.18	4.31	3.25		
Q20b. You had too little time to perform daily tasks					
Always	5.82	4.76	5.29	1.75	0.14
Often	18.29	17.38	17.83		
Sometimes	33.62	38.13	35.88		
Rarely	30.16	31.06	30.61		
Never	12.11	8.67	10.39		
Q20c. You were able to do almost everything you needed to do					
Always	14.69	14.33	14.51	0.35	0.84
Often	38.38	40.96	39.67		
Sometimes	30.96	29.48	30.22		
Rarely	14.48	13.44	13.96		
Never	1.49	1.80	1.65		
Q20d. You were trying to do too many things at once					
Always	11.4	10.94	11.17	0.38	0.82
Often	27.36	26.90	27.13		
Sometimes	33.78	36.71	35.25		
Rarely	19.99	18.72	19.35		
Never	7.47	6.73	7.1		
Q20e. You had more things to do than you could handle					
Always	6.15	6.57	6.36	0.33	0.86
Often	17.87	16.68	17.27		
Sometimes	34.11	32.34	33.22		
Rarely	30.88	33.13	32.01		
Never	10.98	11.29	11.14		
Q20f. You felt too much was expected of you					
Always	7.35	7.06	7.20	0.48	0.75

	Incomplete	Complete	Total	Design-based F	<i>p</i>
Often	16.11	16.59	16.35		
Sometimes	31.12	28.03	29.57		
Rarely	28.83	31.25	30.04		
Never	16.6	17.07	16.84		
Q20g. You had enough time to help others					
Always	10.44	10.43	10.44	0.39	0.81
Often	27.73	29.02	28.38		
Sometimes	43.29	43.14	43.21		
Rarely	16.81	15.11	15.96		
Never	1.73	2.30	2.01		
Q39a. Please indicate whether you think your community spends too much, about the right amount, or too little on each of the following. Would you say...					
Police					
Too much	13.10	14.40	13.75	0.74	0.47
About the right amount	66.57	67.70	67.13		
Too little	20.33	17.90	19.12		
Q39b. Fire					
Too much	2.76	5.17	3.96	5.34	<0.0001
About the right amount	74.60	77.56	76.07		
Too little	22.63	17.27	19.97		
Q39c. Ambulance service					
Too much	2.02	4.24	3.12	3.57	0.03
About the right amount	78.69	79.32	79.00		
Too little	19.29	16.44	17.87		
Q39d. Schools					
Too much	11.09	12.46	11.77	0.40	0.67
About the right amount	55.31	53.60	54.46		
Too little	33.60	33.95	33.77		
Q39e. Library					
Too much	6.57	5.29	5.94	0.71	0.49
About the right amount	72.05	74.37	73.20		
Too little	21.38	20.34	20.87		
Q39f. Parks					
Too much	3.73	3.18	3.45	0.16	0.85
About the right amount	74.06	74.12	74.09		
Too little	22.21	22.70	22.45		
Q39g. Hiking, biking or running trails					
Too much	8.24	6.00	7.13	1.96	0.14
About the right amount	70.21	69.31	69.77		
Too little	21.55	24.68	23.11		
Q39f. Hospitals					
Too much	8.05	8.20	8.12	0.90	0.41
About the right amount	76.48	79.01	77.74		
Too little	15.47	12.79	14.13		
Q39g. Prisons					
Too much	17.25	16.54	16.89	0.87	0.42
About the right amount	62.85	60.52	61.68		
Too little	19.90	22.94	21.42		
Q39h. Mental health facilities					
Too much	2.28	1.86	2.07	0.22	0.81
About the right amount	49.89	51.06	50.48		
Too little	47.83	47.08	47.46		

	Incomplete	Complete	Total	Design-based F	<i>p</i>
Q39i. Rec centers / swimming pools					
Too much	4.41	5.38	4.89	0.35	0.70
About the right amount	73.28	72.21	72.75		
Too little	22.31	22.41	22.36		
Q39j. Childcare					
Too much	5.70	4.95	5.33	0.39	0.67
About the right amount	69.74	68.74	69.24		
Too little	24.56	26.31	25.43		
Q39k. Internet providers					
Too much	16.23	15.35	15.79	0.66	0.52
About the right amount	68.94	67.57	68.26		
Too little	14.83	17.08	15.95		
Q39l. Cable and satellite television providers					
Too much	20.09	19.92	20.01	2.16	0.12
About the right amount	70.76	67.30	69.03		
Too little	9.15	12.78	10.96		
Q41a. How often do you experience each of the following issues related to safety? Would you say...					
I feel safe where I live					
Always	57.42	52.65	55.01	1.08	0.36
Often	34.72	39.90	37.34		
Sometimes	6.42	5.76	6.09		
Rarely	0.63	1.12	0.88		
Never	0.80	0.57	0.69		
Q41b. I avoid places in my town where I do not feel safe					
Always	25.15	21.91	23.51	1.77	0.13
Often	20.81	22.49	21.67		
Sometimes	21.99	18.87	20.41		
Rarely	16.11	16.97	16.55		
Never	15.94	19.75	17.87		
Q41c. I worry about becoming a victim of a crime					
Always	5.85	5.05	5.45	1.74	0.14
Often	7.91	4.88	6.37		
Sometimes	26.28	28.94	27.63		
Rarely	42.80	41.62	42.20		
Never	17.15	19.51	18.34		
Q41d. I worry about someone I care for becoming a victim of a crime					
Always	9.69	8.15	8.91	1.94	0.10
Often	16.88	12.30	14.57		
Sometimes	34.29	38.60	36.47		
Rarely	27.07	27.60	27.34		
Never	12.06	13.35	12.71		
Q41e. I worry about identity theft					
Always	13.17	15.09	14.14	2.25	0.06
Often	22.63	19.75	21.18		
Sometimes	40.71	35.76	38.21		
Rarely	17.39	21.10	19.27		
Never	6.10	8.29	7.21		

	Incomplete	Complete	Total	Design-based F	<i>p</i>
Question Type 2: Incomplete Ordinal Question Stem					
Q9. All in all, do you think that the coronavirus outbreak has been...					
Made a bigger deal than it really is	37.98	38.47	38.22	0.32	0.73
Approached about right	27.65	28.99	28.32		
Made a smaller deal than it really is	34.37	32.54	33.46		
Q34. In the next month, do you think the coronavirus outbreak in Nebraska will...					
Get better	21.17	19.78	20.47	1.50	0.22
Stay about the same	42.78	39.69	41.22		
Get worse	36.05	40.53	38.31		
Q45. Do you think race relations in the United States are ...					
Generally good	43.04	41.84	42.45	0.21	0.65
Generally bad	56.96	58.16	57.55		
Q46. Do you think race relations in the United States are...					
Getting better	16.02	14.51	15.27	0.75	0.47
Stay about the same	28.42	26.76	27.60		
Get worse	55.56	58.73	57.14		
Q47. Do you think race relations in Nebraska are...					
Generally good	72.80	66.40	69.63	6.82	<.0001
Generally bad	27.20	33.60	30.37		
Q48. Do you think race relations in Nebraska are...					
Getting better	16.71	16.46	16.59	3.07	0.05
Stay about the same	58.97	53.48	56.25		
Getting worse	24.33	30.06	27.16		
Q49. Is police brutality in the United States...					
A very serious problem	36.96	29.51	33.21	5.66	<.0001
A somewhat serious problem	32.66	38.47	35.59		
Not a very serious problem	21.22	25.86	23.56		
Not a problem at all	9.16	6.16	7.65		
Q61b. Is your internet service at home...					
Very dependable	36.34	33.07	34.68	0.65	0.62
Mostly dependable	46.18	47.24	46.72		
Somewhat dependable	12.06	13.57	12.83		
A little dependable	3.15	2.85	3.00		
Not dependable at all	2.26	3.26	2.77		
Q61c. Is your internet service at home...					
Very fast	24.25	26.76	25.53	1.21	0.30
Somewhat fast	58.75	57.57	58.15		
Somewhat slow	14.45	11.97	13.19		
Very slow	2.55	3.71	3.14		
Q67. In general, would you describe your political views as...					
Very conservative	8.57	11.31	9.96	1.10	0.36
Conservative	30.64	29.18	29.90		
Moderate	42.54	42.01	42.27		
Liberal	13.35	11.66	12.49		
Very liberal	4.90	5.84	5.38		
Q71. For you, is dropping off outgoing mail like surveys, letters, or greeting cards...					

	Incomplete	Complete	Total	Design-based F	<i>p</i>
Very convenient	51.19	43.64	47.35	2.97	0.01
Somewhat convenient	22.19	25.47	23.86		
Neither convenient nor inconvenient	18.29	20.83	19.58		
Somewhat inconvenient	7.27	7.32	7.29		
Very inconvenient	1.06	2.74	1.92		
Type 3: Incomplete Nominal Question Stem					
Q52. Is the building where you live...					
A mobile home	2.42	2.61	2.51	0.84	0.48
A one-family house detached	77.05	73.12	75.11		
A one-family house attached	4.94	5.31	5.12		
A building with apartments	15.59	18.52	17.03		
A boat, RV, van, etc.	0.00	0.44	0.22		
Q53. Is our home...					
Owned by you b/loan	47.16	44.00	45.54	1.08	0.35
Owned by you free and clear	23.44	27.33	25.42		
Rented	28.26	27.40	27.82		
Occupied without payment of rent	1.15	1.28	1.21		
Q54. Are you...					
Married	61.70	58.60	60.17	1.40	0.23
Unmarried, living with a partner	6.19	8.32	7.24		
Widowed	6.14	5.65	5.90		
Divorced	8.21	10.92	9.54		
Separated	0.95	0.62	0.79		
Single, never married	16.80	15.89	16.35		
Q55. Do you consider yourself to be...					
Heterosexual or straight	92.94	93.91	93.43	0.61	0.52
Gay or lesbian	1.96	1.15	1.55		
Bisexual	5.09	4.94	5.01		
Q63. Do you have...					
A smartphone	78.73	81.30	80.05	4.52	0.003
Both a cell phone and a smartphone	7.09	3.31	5.15		
A cell phone, but not a smartphone	12.77	13.14	12.96		
None	1.42	2.25	1.84		
Q72. Did you complete this questionnaire...					
At home	94.39	93.06	93.71	1.10	0.35
At work or school	4.87	5.74	5.31		
At a coffee shop, café, car, or restaurant	0.00	0.50	0.26		
In a car, train, or bus,	0.21	0.27	0.24		
Other (specify):	0.53	0.43	0.48		

Table A.6 Summary of Hypotheses and Results

Item	Overall	Any Expected Differences for...			Results?
		Incomplete conversational	Incomplete ordinal	Incomplete nominal	
Nonresponse	Incomplete question stems more burdensome and have higher item nonresponse rates (H1)	Incomplete same as complete (H1a)	Incomplete lower than complete (H1b)	Incomplete same as complete (H1c)	No difference between incomplete and complete (H1 and H1b not supported; H1a, H1c supported)
Response time	Incomplete question stems more burdensome and have longer response time (H2)	--	Incomplete shorter than complete (H2b)	Incomplete shorter than complete (H2c)	Incomplete longer than complete (H2 supported; H2b and H2c not supported)
Select first	Attitude questions: Incomplete have higher rates of selecting first answers; Factual questions: No difference for incomplete and complete (H3)	--	--	--	No difference between incomplete and complete (H3 not supported)
Select last	Attitude questions: Lower rates of selecting last answers for incomplete questions; Factual questions: No different for incomplete and complete (H4)	--	--	--	Lower rates of selecting last answers for incomplete than complete for attitudinal questions (H4 supported)

Note: -- indicates that the pattern for the specific type of incomplete question stem is consistent with the overall expectation.