

More Clarification, Less Item Nonresponse in Establishment Surveys? A Split-Ballot Experiment

Benjamin Kufner

Institute for Employment Research (IAB)
Nuremberg, Germany

Joseph W. Sakshaug

Institute for Employment Research (IAB)
Nuremberg, Germany,

University of Mannheim, Germany and
Ludwig Maximilian University of Munich, Germany

Stefan Zins

Institute for Employment Research (IAB)
Nuremberg, Germany

The IAB Job Vacancy Survey of the German Institute for Employment Research collects detailed information on job search and vacancy durations for an establishment's last successful hiring process. The duration questions themselves are burdensome for respondents to answer as they ask for precise dates of the earliest possible hiring for the vacancy, the start of the personnel search, and the decision to hire the selected applicant. Consequently, the nonresponse rates for these items have been relatively high over the years (up to 21 percent). In an effort to reduce item nonresponse, a split-ballot experiment was conducted to test the strategy of providing additional clarifying information and examples to assist respondents in answering the date questions. The results revealed a backfiring effect. Although there was evidence that respondents read the additional clarifying information, this led to even more item nonresponse and lower data quality compared to the control group. Additionally, we observed a negative spillover effect with regard to item nonresponse on a subsequent (non-treated) question. We conclude this article by discussing possible causes of these results and suggestions for further research.

Keywords: questionnaire design; cognitive response processes; business survey; missing data

1 Introduction

Item nonresponse is a common problem in surveys. The failure to collect responses to all survey items reduces the analytic sample size and consequently leads to a reduction in estimation precision and statistical power. It can also lead to biased estimates if the missing values are not missing completely at random (Little & Rubin, 2019). Hence, item nonresponse is an important and impactful source of survey error (De Leeuw, Hox, & Huisman, 2003). It is often associated with response burden and can be driven by several factors, including questionnaire design (e.g. Messmer & Seymour, 1982; Zuell, Menold, & Körber, 2015), the cognitive response process (Tourangeau, Rips, & Rasinski, 2000), question wording (e.g. Lenzner, Kaczmirek, & Lenzner, 2010), and respondent instructions (e.g. Al Baghal & Lynn, 2015; Smyth, Dillman, Christian, & McBride, 2009), which are

well-documented in the experimental household survey literature.

However, experimental evidence on the impact of questionnaire design on item nonresponse in establishment surveys is rare (Bavdaž et al., 2020). Thus, it is often unclear whether design decisions intended to reduce item nonresponse are effective or not, or do more harm than good. As an exception, Ott, McGovern, and Sirkis (2016) experimented with question ordering and showed that asking about personal characteristics at the beginning of an establishment survey questionnaire reduced item nonresponse compared to asking at the end. In another experiment, O'Brien and Levin (2007) showed that using answer prompts reduced item nonresponse in a web survey of establishments. A notable gap in the empirical literature are experiments which test the impact of providing additional respondent instructions and clarifications on item nonresponse in establishment surveys.

In this article, we test whether providing additional clarifying information reduces item nonresponse to three vacancy duration questions affected by high item nonresponse rates in the IAB Job Vacancy Survey. Specifically, we conducted a

Contact information: Benjamin Kufner, Regensburger Str. 100, 90478 Nürnberg Germany (E-Mail: benjamin.kuefner@iab.de)

split-ballot experiment by supplementing the questions with clarifying examples and easy-to-look-up events to aid respondents in providing an adequate response. We tested the hypothesis that providing this additional information would reduce item nonresponse and improve data quality.

2 Background

2.1 Item Nonresponse in the IAB Job Vacancy Survey (IAB-JVS)

One specific question topic that can be especially burdensome for establishments to answer are those related to job vacancies and vacancy durations, which may require look-up of internal records to identify precise temporal information. Several job vacancy surveys are conducted worldwide to collect this information (e.g. Statistics Canada, 2020; U.S. Bureau of Labor Statistics, 2020). One example is the IAB Job Vacancy Survey (IAB-JVS) in Germany, which is a large, annual, and nationally-representative establishment survey that collects extensive information about job vacancies, job flows, and search and recruitment processes (Bossler, Gürtzgen, Kubis, Kufner, & Lochner, 2020). It is carried out using a concurrent mixed-mode (mail and web) design. The survey includes, among other items, three questions that ask for specific dates concerning the last successful search and recruitment process. In particular, these questions ask for the earliest possible hiring date for the posted position (“At what date should this position be filled at the earliest?”), the date the search started (“When did you start searching for this vacancy?”), and the date the decision was made to hire the applicant (“When did you decide to hire this applicant?”).

These specific date variables are frequently used by researchers in conjunction with variables about the recruitment process and the employment biographies of the hired applicant to conduct unique analyses on the structure of the labor demand, the efficiency of search and recruitment methods, and the quality of the resulting job match (e.g. Gürtzgen, Kubis, & Kufner, 2019; Gürtzgen & Moczall, 2020; Rebien, 2019). However, these questions have been facing a disproportionately high share of item nonresponse in recent years (see Table 1), ranging regularly between 12 and 21 percent, with a slightly higher share for the search start date than the other two date items, thus compromising the quality and utility of these data.

2.2 Possible Causes of Item Nonresponse for IAB-JVS Vacancy Duration Questions

To get an idea of what might cause item nonresponse in the vacancy duration questions, it is useful to consider the cognitive model of the survey response process (see Figure 1, Row 1) proposed by Tourangeau (1984) and Tourangeau et al. (2000) for household surveys and further adaptations by

Table 1
Item Nonresponse Rates (in %) for Vacancy Duration Questions in the 2015–2018 IAB Job Vacancy Survey

Year	Earliest hiring date	Start search date	Applicant decision date
2015	17.64	21.01	17.08
2016	13.55	16.99	13.31
2017	13.67	16.86	13.30
2018	12.72	16.03	12.44

Bavdaž (2010) and Willimack and Nichols (2010) for establishment surveys. They divide the survey response process into four main components: comprehension, retrieval, judgment and response. Comprehension is defined as “identifying the information sought” (Tourangeau et al., 2000, p. 9) from the question and the accompanying instructions. Retrieval refers to the process of gathering the requested information, which for establishments is strongly linked to the availability of and access to the requested data in their business record systems (e.g. Bavdaž, 2010; Lorenc, 2007; Willimack & Nichols, 2010). The judgment component encompasses judgments and estimations based on the retrieved data and evaluation of the “adequacy of a response” (Willimack & Nichols, 2010, p. 14). Finally, the response component addresses the processes of matching a judgment to an answer field and editing the answer accordingly (Tourangeau et al., 2000, pp. 13–14).

Using the survey response model as a basis, it is possible to identify potential problems in the components that might prevent establishments from answering the vacancy duration questions in the IAB-JVS (see Figure 1, Row 2). With respect to comprehension, establishments might find the concepts of earliest hiring date, search start date, and the applicant decision date to be rather vague. Because the questions are standardized across all establishment types and hiring processes, how establishments comprehend the questions may not correspond precisely to their own recruitment processes. Accordingly, if establishments are unable to map the question onto their specific hiring case and link the survey question to their “business reality” (Bavdaž, 2010), then item nonresponse could arise. In addition, the concepts may be perceived as ambiguous or vague to respondents. For example, establishments might be unsure whether the earliest possible hiring date refers to the first day when the position was vacant after a dismissal, when a new project was started, or when the funding for the new position became available. Similarly, the applicant decision date could be interpreted in multiple ways: when the supervisor agreed to hire the candidate or when the budget manager agreed to the hire. Another potential problem that could arise is due to presupposition

of the questions. For instance, the start search date question assumes that the company has undertaken a formal search for applicants. However, this underlying assumption may be violated if no search for an applicant was necessary (e.g. the position was filled via an unsolicited application without a job posting). Vague and ambiguous concepts and presupposition are common problems that affect respondents' ability to answer survey questions (Bavdaž, 2010; Haraldsen, 2013; Tourangeau et al., 2000; Willimack & Nichols, 2010).

Retrieval of the requested search and recruiting dates is highly dependent on the availability of this information in the establishment's record systems. While professional recruitment software may facilitate retrieval or deduction of the requested dates, establishments without such software would need to search alternative sources, such as proceedings or internal communications (e.g. emails) with applicants, supervisors or other departments, which is a more burdensome proposition that respondents may view as not worth the effort (Willimack & Nichols, 2010, pp. 13–14). If respondents cannot find the relevant dates in their business records, then they would have to rely on their memory or that of colleagues. Such dates could be difficult to recall if the events in question occurred many weeks or months ago or were not highly salient or distinguishable from other human resource activities. In this situation providing respondents with "temporal landmarks" (Tourangeau et al., 2000, pp. 115–117) could guide respondents in their memory process to help them remember (or narrow down) the possible dates.

Possible reasons for item nonresponse to the vacancy duration questions could also occur during the judgment step. Related to the ambiguity of key concepts, respondents may retrieve multiple dates that are considered applicable to the questions. If no additional distinguishing information is provided in the question or accompanying instructions, then it may be difficult for respondents to decide which is the "most appropriate" date (Willimack & Nichols, 2010, pp. 14–15), thus creating a burden of choice that may lead to no answer. If no readily-available date information is found in the business records, then respondents may have to estimate the dates based on vague information or informal communications with colleagues (Lorenc, 2007). This can also be a burdensome and error-prone task that respondents may not be willing to engage with if the respondent was not directly involved in the hiring process, or the recruiting process occurred long ago and the information base is poor (Tourangeau et al., 2000).

Fewer causes of item nonresponse are envisaged during the response step. As dates can be entered either by entering the date into the response field or by selecting the date from a pop-up calendar, the completion process seems less prone to problems of nonresponse. In addition, linking the retrieved dates to the response field seems also straightforward and poses little risk of confusion that may prevent a response.

However, a potential problem could arise if respondents retrieve or remember only a specific week or month rather than the exact date of the event. As an exact date is required in the IAB-JVS, respondents cannot enter their imprecise answer or a date range into the response field, which may lead to either guessing the exact date or not providing a response at all.

2.3 Providing Clarifying Information to Reduce Item Nonresponse

To address some of the problems identified above, survey guidelines recommend supplementing complicated questions with clarifying details and examples. For instance, Redline (2011) asserts that respondents could be assisted by providing clarifying information, such as examples or further instructions, to reduce question vagueness, ambiguity, and mapping discrepancies. In their guidelines for establishment questionnaires, Morrison, Dillman, and Christian (2010) highlights the importance of instructional information: "Particularly in establishment surveys, the instructions are often very important for conveying the correct specifications or intent of the question..." (Morrison et al., 2010, p. 64). In addition, the Forms Design Standards Manual of the Australian Bureau of Statistics (2010) suggests to use examples to assist respondents for demanding questions, such as open-ended questions. In establishment surveys, it is common to provide these clarification instructions in separate materials or help pages, but there is a risk that respondents do not recognize or consider this additional information when answering the relevant survey questions (Haraldsen, 2013; Morrison et al., 2010).

In the case of the IAB-JVS, providing clarifying information and examples of milestone events in a general hiring process could reduce item nonresponse by improving respondents' comprehension of the vacancy duration questions and make the concepts seem less vague or ambiguous by guiding respondents toward the intended interpretation of the key concepts. Providing additional information is, however, unlikely to address all potential comprehension problems (e.g. presupposition). With regard to retrieval problems, providing additional clarifying information and examples of milestone events could help respondents find the corresponding event date information in their record systems, or improve their search in other documents (e.g. emails). Even for establishments without a documentation of these vacancy dates, the provided examples could serve as "temporal landmarks" and assist respondents in remembering the hiring process dates in more detail. Additionally, a positive impact of the clarifying information may be expected during the judgment step, where it aids respondents in selecting the most adequate date out of several applicable options and hence reduces the burden of choice. Finally, the response stage could be facilitated if the clarifying information leads to the identification of exact dates, which fit the required format of the response field.

	Comprehension	Retrieval	Judgment	Response
Process	<ul style="list-style-type: none"> • Interpretation of the Question • Identifying the information sought 	<ul style="list-style-type: none"> • Checking business records • Remembering recruiting process 	<ul style="list-style-type: none"> • Estimations based on retrieved data • Evaluation of adequacy of a response 	<ul style="list-style-type: none"> • Mapping response and answer field • Editing response
Possible causes of item non-response	<ul style="list-style-type: none"> • Vagueness of questions • Ambiguity of questions • Presupposition of questions 	<ul style="list-style-type: none"> • Availability of vacancy dates in business records • Events difficult to remember 	<ul style="list-style-type: none"> • Burden of choice between adequate dates • Estimation based on poor information 	<ul style="list-style-type: none"> • Provision of exact dates

Figure 1. Cognitive Model of the Survey Response Process (Tourangeau, Rips, & Rasinski, 2000, p. 8; Bavdaž, 2010, p. 83) and Possible Causes of Item Nonresponse for Vacancy Duration Questions

3 Data & Methods

3.1 Experiment

Based on the previously discussed causes of item nonresponse in the IAB-JVS and clarifying information as a possible remedy for these problems, we experimentally added additional and more specific instructions to each question based on the aforementioned questionnaire design recommendations. These additional instructions included concrete examples of process dates for specific milestone events that the establishment could use and adapt to their own hiring situation. Expert interviews were conducted with substantive researchers and survey methodologists to identify two exemplary milestone dates for each of the asked questions. These exemplifying events are not a comprehensive list of all possible events, but should give respondents an impression of the intended meaning of the questions.

For the earliest hiring date question, a random half of the respondents were provided the following additional information:

“Here you could, for example, enter the following events:

- *Date of a possible project start, in which the new employee should participate*
- *For replacements: Day on which the position is vacant”.*

For the start search date question, the treatment group was presented with the following examples:

“Here you could, for example, enter the following events:

- *Date of publication of the job advertisement (e.g. on the homepage, a newspaper or an online job market)*
- *Date of public posting”.*

Finally, the applicant decision date question included the following additional information for the treatment respondents:

“Here you could, for example, enter the following events:

- *Date of final approval by the supervisor*
- *Date of the final interview”.*

Screenshots of the treated questions, including translations, are provided in Figure 2.

As proposed by Couper (2008), the additional information was displayed in italics between the question and the answer box. As it lies within the reading logic (up to down), the information is likely to be noticed and read by respondents. Hence, we overcome the limitation of separate instruction pages (Morrison et al., 2010). This is checked by examining item durations. We also analyze one additional item from

the IAB-JVS, which asks for the date the employment relationship began (“When did this employment relationship begin?”). This item was not experimentally treated because it is less affected by item nonresponse than the three experimental items. However, since the employment relationship item directly follows (and has a similar appearance and format) to the three treated items, the treatment effects (if any) are likely to carry over to this item as well. Hence, we investigated a possible spillover effect on this non-treated item. These four questions were not preceded by an explicit introduction explaining reasons or motivation for asking about search and recruiting events.

3.2 Survey Design

As previously noted, the three treated questions are essential questions of the IAB-JVS survey. These questions belong to the module on the establishment’s last successful hiring. Only establishments who reported a successful hiring within the last year are asked to complete this module (“Please think of the last hire of a new employee into a position subject to social security contributions in the past 12 months. If more than one person was hired at the same time, please choose the person whose last name comes first in the alphabet.”). Hence, these questions are only presented to a subset of eligible respondents. To avoid tinkering with the main production survey, the experiment was conducted within a separate survey that ran parallel to the main survey in 2019.¹ In contrast to the main survey, which uses a concurrent mixed-mode (mail/web) design, the experimental survey (including both treatment and control groups) was implemented entirely online. Besides this difference, the survey design was very similar to that of the main IAB-JVS survey. The questionnaire and the organization of the fieldwork (corresponding invitation and reminder letters) were the same (for more design details, see Table 2).

The experimental survey used a similar stratified random sampling design as the main survey with industry and establishment size as stratification variables. The sample size for the survey was 31,905. The sample size was chosen according to the projected unit response rate and share of eligible respondents. This calculation led to an expected realized sample of at least 1,750 respondents eligible for the experiment, i.e. a planned eligible respondent sample size of 875 for both treatment and control groups. The planned number of eligible respondents was estimated from a power calculation of a two-sample *t*-test to detect an effect size of 3 percent on a confidence level of 95 percent.

To improve the efficiency of the sampling design and thus the power of the planned statistical tests, an optimal sample allocation was used. The allocation was optimized towards the historical distribution of item nonresponse in the treated questions. To solve the allocation problem, we applied the method described by Friedrich, Münnich, and Rupp (2018)

Table 2
Summary of Survey Design

Mailing Date of Invitation Letter	30 th September 2019
Mailing Date of Reminder	20 th November 2019
Survey Mode	Online
Full Sample	31,905
Control	15,939
Treatment	15,966
Net Sample (Respondents)	4,414
Control	2,225
Treatment	2,189
Unit Response Rate (in %)	13.83
Eligible Respondents	1,683
Control	844
Treatment	839

to the empirical distributions of the item nonresponse indicators from the previous year’s (2018) survey. However, the gain in efficiency due to the optimal allocation over a proportional (to stratum size) allocation was minor, with an average reduction of the expected design effect for the item response indicators of only 0.86 percent. Each sample stratum was split into two approximately equal-sized sets defining the experimental and control groups.

The unit response rate was 13.83 percent. Out of 4,414 responding establishments, a total of 1,683 were eligible for the experiment, with 844 responding from the control group and 839 from the treatment group. There are no significant differences between control and treatment cases with respect to establishment characteristics (see Table 3). Hence, the experimental design worked as intended. The item nonresponse rate is calculated by dividing the number of establishments that did not answer the respective question and the number of eligible establishments.² As a “don’t know” or refusal option was not provided, respondents were forced to answer or skip a question. Each question was displayed on a separate screen and time stamps were used to record the time taken to answer

¹This analysis is based on a preliminary dataset. The final data will be accessible from mid-2022 at the Research Data Center of the Federal Employment Agency in Germany. The final data set is adjusted for respondents who did not answer items relevant to Eurostat. This exclusion has no substantial impact on the results presented.

²A small number of break-offs ($N = 18$) occurred between the branch question, which determines whether the respondent is eligible to answer the treated questions, and the last analyzed question. These are counted as item nonrespondents. Additional sensitivity checks (not reported) showed that excluding these break-offs from the item nonresponse analysis did not substantially change the results.







Treatment group	Control group
<p>35. Zu welchem Termin sollte diese Stelle frühestens besetzt werden?</p> <p><i>Hier könnten Sie folgende Zeitpunkte beispielsweise eintragen:</i></p> <ul style="list-style-type: none"> • Datum eines möglichen Projektstartes, bei dem der neue Mitarbeiter mitwirken soll • Bei Ersatz Einstellungen: Tag, an dem die Stelle erstmalig unbesetzt ist <div data-bbox="172 607 799 703"> <input type="text"/>  <p>Format: tt.mm.jjjj</p> </div> <p><i>Translation:</i> At what date should this position be filled at the earliest? Here you could, for example, enter the following events: Date of a possible project start, in which the new employee should participate. For replacements: Day on which the position is vacant</p>	<p>35. Zu welchem Termin sollte diese Stelle frühestens besetzt werden?</p> <div data-bbox="831 539 1458 636"> <input type="text"/>  <p>Format: tt.mm.jjjj</p> </div> <div data-bbox="831 674 903 707"> <p>Zurück</p> </div> <p><i>Translation:</i> At what date should this position be filled at the earliest?</p>
<p>36. Wann haben Sie mit der Personalsuche für diese Stelle begonnen?</p> <p><i>Hier könnten Sie folgende Zeitpunkte beispielsweise eintragen:</i></p> <ul style="list-style-type: none"> • Datum der Veröffentlichung der Stellenanzeige (z.B. auf der Homepage, einer Zeitung oder einer Online-Stellenbörse) • Datum des öffentlichen Aushangs <div data-bbox="172 1070 799 1167"> <input type="text"/>  <p>Format: tt.mm.jjjj</p> </div> <p><i>Translation:</i> When did you start searching for this vacancy? Here you could, for example, enter the following events: Date of publication of the job advertisement (e.g. on the homepage, a newspaper or an online job market); Date of public posting</p>	<p>36. Wann haben Sie mit der Personalsuche für diese Stelle begonnen?</p> <div data-bbox="831 1003 1458 1099"> <input type="text"/>  <p>Format: tt.mm.jjjj</p> </div> <div data-bbox="831 1137 903 1171"> <p>Zurück</p> </div> <p><i>Translation:</i> When did you start searching for this vacancy?</p>
<p>37. Wann haben Sie sich für diesen Bewerber entschieden?</p> <p><i>Hier könnten Sie folgende Zeitpunkte beispielsweise eintragen:</i></p> <ul style="list-style-type: none"> • Datum der finalen Zustimmung durch den Vorgesetzten • Datum des finalen Vorstellungsgesprächs <div data-bbox="172 1532 799 1628"> <input type="text"/>  <p>Format: tt.mm.jjjj</p> </div> <p><i>Translation:</i> When did you decide to hire this applicant? Here you could, for example, enter the following events: Date of final approval by the supervisor; Date of the final interview</p>	<p>37. Wann haben Sie sich für diesen Bewerber entschieden?</p> <div data-bbox="831 1460 1458 1556"> <input type="text"/>  <p>Format: tt.mm.jjjj</p> </div> <div data-bbox="831 1594 903 1628"> <p>Zurück</p> </div> <p><i>Translation:</i> When did you decide to hire this applicant?</p>

Figure 2. Screenshots and Translations of the Vacancy Duration Questions, by Treatment and Control Groups

Table 3
Experimental Allocation of Respondents to Control and Treatment Groups, by Establishment Characteristics

Variable	Control	Treatment
Establishment Size		
1–9	544	528
10–19	127	142
20–49	116	118
50–249	50	49
≥250	7	2
Industry		
Primary Sector	25	23
Secondary Sector	220	224
Logistics and Retail	163	176
Other Services	422	403
Public Sector	14	13
Region		
West Germany	687	677
East Germany	157	162
Collective Agreement		
Yes	548	525
No	290	303
Item Missing	6	11
Total	844	839

χ^2 -tests showed no significant differences between the control and treatment groups regarding the displayed descriptive statistics

or skip a question. Outliers that took more than 15 minutes to answer or skip the question ($N = 6$) are excluded from the item duration analysis. All statistical analyses, including two-sample t -tests, account for the survey design.³

4 Results

4.1 Item Duration

First, we assess the item durations to check whether respondents in the treatment group likely read the additional clarifying text. Figure 3 shows the average item durations for all treated questions by treatment and control group. Additionally, it distinguishes between all eligible participants (upper panel), item respondents (middle panel), and item nonrespondents (lower panel).

Starting with all eligible participants, it is apparent for all three questions that establishments in the treatment group took on average significantly longer to proceed to the next question compared to the control group. This difference is largest for the first question (earliest hiring date: 56 vs. 30 seconds; $p = 0.000$) followed by the second question (search start date: 29 vs. 23 seconds; $p = 0.000$) and third question

(applicant decision date: 31 vs. 26 seconds; $p = 0.063$). The pattern for item respondents is very similar to the one for all eligible participants: the treatment group needed significantly more time than the control group to respond to each of the three items. Hence, we conclude that item respondents in the treatment group likely read the additional information provided.

Item nonrespondents behave in a similar way. First, as expected, they took less time to go forward to the next question compared to item respondents. Secondly, the longer item duration for the treatment group is significant for all three items. Hence, we conclude that the item nonrespondents also likely read the additional information in the treated items.

4.2 Item Nonresponse

Next, we analyze the main outcome: item nonresponse. Specifically, we examine the nonresponse rates for the three treated questions individually, along with two summary nonresponse indicators: whether a nonresponse occurred in at least one of the treated questions and whether nonresponse occurred in all three treated questions for a given establishment. Figure 4 presents the item nonresponse rate for the three questions and the rates of the two summary indicators.

The results show a consistent pattern: contrary to expectations, item nonresponse is higher in the treatment group than in the control group for all three items. The item nonresponse rates in the treatment group are 15.3 percent, 19.0 percent, and 19.9 percent for the earliest hiring date, start search date, and applicant decision date items, respectively, compared to the respective control group values of 12.4 percent, 15.0 percent, and 13.6 percent. Thus, the item nonresponse rates range from 2.9 to 6.6 percentage points higher in the treatment group compared to the control group. These differences are statistically significant at the 10 percent level for the earliest hiring date and at the 5 percent level for the start search date and the applicant decision date items. In addition, both summary indicators show a negative effect of the treatment on the item response rate. The rate for the indicator of at least one missing item in the treatment group is 6.9 percentage points above the control group and is statistically significant at the 5 percent level. The rate for the all-missing indicator in the treatment group lies 2.7 percentage points above the control group and is significant at the 10 percent level. In summary, providing the exemplifying information to respondents actually led to more item nonresponse compared to not providing this additional information.⁴

³Weighting to account for unequal inclusion probabilities and unit nonresponse does not change the study conclusions (results not shown).

⁴The negative treatment effect was also observed for several establishment subgroups, including establishment size, industry, region, and collective agreement (results not shown).

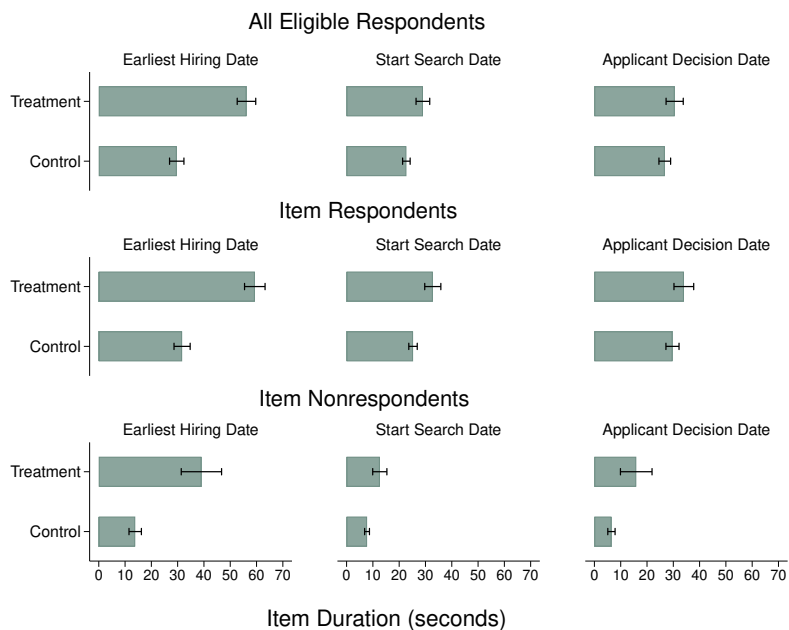


Figure 3. Average Item Durations (in Seconds) for All Eligible Participants (Upper Panel), Item Respondents (Middle Panel), and Item Nonrespondents (Lower Panel), by Item and Treatment and Control Groups.

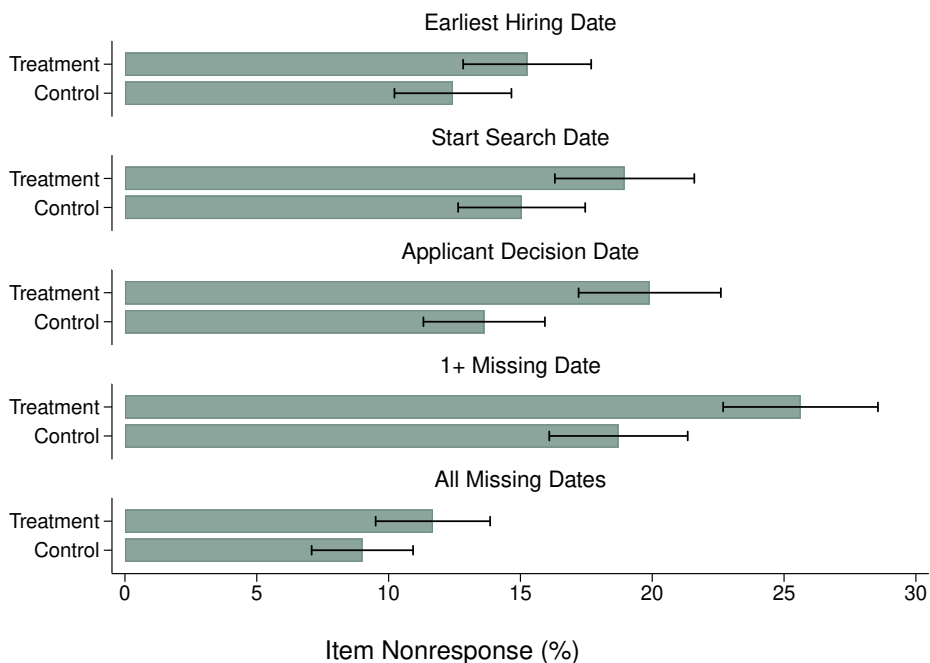


Figure 4. Rates of Item Nonresponse (%) by Treatment and Control Groups by Item, At least One Missing Item, and All Missing Items

4.3 Data Quality

Here, we assess the impact of the treatment on data quality. Although it is not possible to assess the accuracy of the dates provided by the responding establishments due to the lack of validation data, it is possible to assess the effects of the treatment on other indicators of data quality.

The indicator we use is the share of implausible date values defined as dates that occur after the survey completion date. Such dates are by definition incorrect, as the establishments are asked to report on the last completed hiring process. This quality indicator is evaluated for the second and third items only (i.e. the search start date and the applicant decision date). It is not applicable to the first item, as it is plausible that the earliest possible hiring date was originally planned for a future date, but these plans were adjusted later. As shown in Figure 5, the share of implausible future dates in the treatment group is 1.5 percentage points higher for the start search item and 2.5 percentage points higher for the applicant decision item compared to the control group. Both differences are statistically significant at the 5 percent level. The negative effect of the treatment is especially evident for the summary indicator of whether at least one answered date lies in the future, where the treatment group is 3.3 percentage points higher than the control group and is statistically significant. No statistically significant effect could be found for the summary indicator of whether all answered dates were implausible between the treatment (0.9 percent) and control groups (0.4 percent). Thus, we conclude that the quality of the responses is poorer when the additional clarifying information is presented to respondents.

4.4 Spillover Effects

Lastly, we examine the possibility of a spillover effect of the treatment on the first non-treated item directly following the three experimental items. This question asks for the date the employment relationship began. As it also asks for a specific date, it has a similar appearance and format as the previous three treated items.

Figure 6 summarizes the relevant outcomes (item duration, item nonresponse, implausible values) for this particular item. Since the question was identical in both control and treatment groups, there is, as expected, no significant difference in response time between the two groups. However, the negative effect of the treatment on item nonresponse observed for the previous three treated items is also observed for this non-treated item. The share of item nonresponse in the control group (9.2 percent) is significantly lower than the share of item nonresponse in the treatment group (14.1 percent). Thus, the effect of the treatment is carried over to the non-treated item with the same question format. We do not find a spillover effect on data quality, as there is no significant difference in the share of implausible values between the two

groups. For further follow-up items, no spillover effect was found (results not shown).

5 Discussion

Contrary to expectations based on questionnaire design recommendations (e.g. Australian Bureau of Statistics, 2010; Morrison et al., 2010; Redline, 2011) and the cognitive model of survey response (Bavdaž, 2010; Tourangeau et al., 2000; Willimack & Nichols, 2010), we found that providing additional clarifying information and examples to assist establishments in answering temporal questions related to the process of hiring their most recent employee did not reduce item nonresponse. Instead, this strategy led to a significant increase in item nonresponse. Moreover, there was evidence that the additional information had a negative effect on data quality, as it led to more implausible answers. We also found a spillover effect where the negative effect of providing the additional information on the item nonresponse rate carried over to an adjacent non-treated question. These results clearly rejected our hypothesis that providing the additional clarifying information would overcome possible response problems, reduce item nonresponse, and improve data quality in the IAB-Job Vacancy Survey.

There are at least four plausible explanations for this unexpected finding. First, the response burden and complexity of these questions may have actually increased as more information was presented and needed to be absorbed. Respondents had to read more text and may have felt pressured to look up their records and provide an exact date rather than surmise a response. The additional pressure of providing exact dates may have led respondents to skip the questions entirely. Second, by presenting the additional information, the importance of these questions was implicitly highlighted. Establishments that were unsure about the exact dates of the asked events may have therefore preferred not to answer the question as opposed to giving an uncertain answer. Third, offering examples of two process dates, which are not mutually exclusive, such as the date of publication of the job advertisement on the company homepage and the date of public posting, may have increased the uncertainty of how to answer the questions instead of reducing it. If both of the examples applied to the establishment and the corresponding dates differed slightly, then respondents may not have known which of them they should use. As a result of this confusion, respondents might have simply skipped the item rather than venture a guess. Additionally, as the list of provided examples was not an exhaustive list of all possible scenarios, respondents who did not see an example of their hiring situation may have believed the question(s) did not apply to them. This might have contributed to the higher item nonresponse.

The IAB-JVS vacancy duration questions could benefit from extensive qualitative research to identify the key reasons for item nonresponse. Such an analysis could clarify

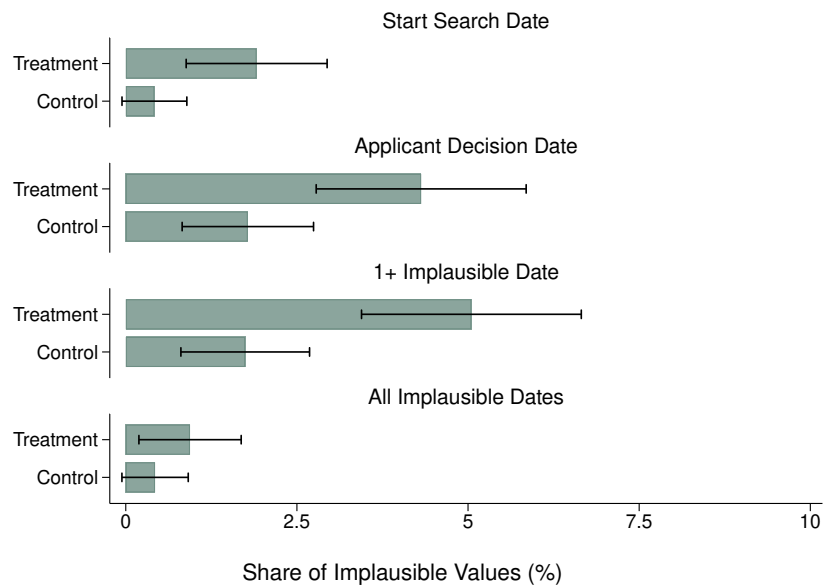


Figure 5. Share of Implausible Values (%) by Treatment and Control Groups by Item, At least One Implausible Date, and All Implausible Dates

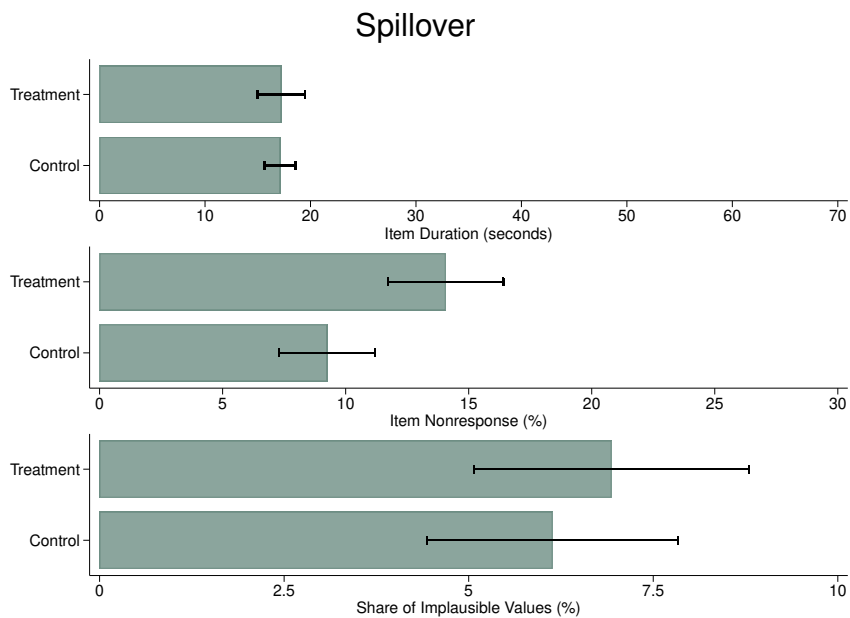


Figure 6. Item Duration (in Seconds), Item Nonresponse (%), and Share of Implausible Values (%) for the First Non-Treated Item Following the Treated Items, by Treatment and Control Groups

which component of the cognitive response model is most problematic for respondents and the main contributor of the high item nonresponse for these items. Another contribution of qualitative research could be to gain more insights into the search and recruiting processes of establishments. This information should be used to evaluate the applicability of the questions themselves. If these questions do not apply to the hiring processes in some establishments, then respondents may refuse to answer them rather than give an inaccurate answer, because they do not want to risk providing misleading information about their processes. By taking these qualitative insights into account new questions or clarifying instructions could be developed for collecting search and recruiting durations. Finally, split-ballot experiments, like the one presented here, could show whether these new questions or additional instructions perform better with respect to item nonresponse and data quality.

More generally, future survey experiments, especially those concerning establishment questionnaire design, would benefit from a mixed-method approach, combining experimental evidence with qualitative research to gain insights into the impact of the treatment on response burden and the entire response process. The contribution of qualitative research could be twofold: (1) qualitative pretests could be used to adjust the treatment closer to the needs of the respondent; and (2) qualitative debriefings with item respondents and nonrespondents could shed light on what, in particular, drives the treatment effect (or lack thereof). Specifically, such an approach would enable researchers to compare differences in the cognitive response process for establishments in the control and treatment groups and hence identify the model component(s) which are most affected by the treatment.

In conclusion, this case study showed that providing establishments with additional clarifying information and examples did not reduce, and rather exacerbated, item nonresponse to a set of job vacancy duration questions. Although the rationale for providing additional clarifying details might be well-justified for complex establishment surveys and in line with questionnaire design recommendations, the implementation, as our study showed, can potentially backfire because of increased response burden or other unintended effects. However, without an experiment, such a backfiring effect may not be exposed. Thus, we encourage similar experiments in real-world establishment survey settings where design modifications are considered to facilitate item response. This would also address the notable gap in the empirical literature on questionnaire design effects in establishment surveys.

Acknowledgements

The Institute for Employment Research supported this work. We are grateful to Economix Research & Consulting

for implementing the experiment. We thank Mario Bossler, Benjamin Börschlein, Nicole Gürtzgen, Alexander Kubis, and Ger Snijkers for helpful comments and suggestions. We are grateful to Vinzenz Pyka for excellent research assistance.

References

- Al Baghal, T., & Lynn, P. (2015). Using motivational statements in web-instrument design to reduce item-missing rates in a mixed-mode context. *Public Opinion Quarterly*, 79(2), 568–579.
- Australian Bureau of Statistics. (2010). *Forms design standards manual*. Canberra: Australian Bureau of Statistics.
- Bavdaž, M. (2010). The multidimensional integral business survey response model. *Survey Methodology*, 36(1), 81–93.
- Bavdaž, M., Snijkers, G., Sakshaug, J. W., Brand, T., Haraldsen, G., Kurban, B., & Willimack, D. K. (2020). Business data collection methodology: Current state and future outlook. *Statistical Journal of the IAOS*, 36(3), 741–756.
- Bossler, M., Gürtzgen, N., Kubis, A., Kufner, B., & Lochner, B. (2020). The IAB Job Vacancy Survey: Design and research potential. *Journal for Labour Market Research*, 54(13), 1–12.
- Couper, M. P. (2008). *Designing effective web surveys*. New York, USA: Cambridge University Press.
- De Leeuw, E. D., Hox, J. J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, 19(2), 153–176.
- Friedrich, U., Münnich, R., & Rupp, M. (2018). Multivariate optimal allocation with box-constraints. *Austrian Journal of Statistics*, 47(2), 33–52.
- Gürtzgen, N., Kubis, A., & Kufner, B. (2019). *Bei befristeten Einstellungen wenden die Betriebe weniger Mittel auf*. IAB-Kurzbericht, 17/2019, Nürnberg.
- Gürtzgen, N., & Moczall, A. (2020). Sequential versus non-sequential search among German employers—evidence from a job vacancy survey. *Applied Economics Letters*, 27(11), 873–879.
- Haraldsen, G. (2013). Questionnaire communication in business surveys. In G. Snijkers, G. Haraldsen, J. Jones, & D. Willimack (Eds.), *Designing and conducting business surveys*. (pp. 303–357). Hoboken, NJ: John Wiley & Sons, Inc.
- Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied cognitive psychology*, 24(7), 1003–1020.
- Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

- Lorenc, B. (2007). Using the theory of socially distributed cognition to study the establishment survey response process. In *Proceedings of the third international conference on establishment surveys* (pp. 881–891). American Statistical Association. Montreal, Canada.
- Messmer, D. J., & Seymour, D. T. (1982). The effects of branching on item nonresponse. *Public Opinion Quarterly*, 46(2), 270–277.
- Morrison, R. L., Dillman, D. A., & Christian, L. M. (2010). Questionnaire design guidelines for establishment surveys. *Journal of Official Statistics*, 26(1), 43–85.
- O'Brien, J., & Levin, K. (2007). Web design issues in a business establishment panel survey. In *Proceedings of the third international conference on establishment surveys* (pp. 1075–1078). American Statistical Association. Montreal, Canada.
- Ott, K., McGovern, P., & Sirkis, R. (2016). Using analysis of field test results to evaluate questionnaire performance. In *Proceedings of the fifth international conference on establishment surveys*. American Statistical Association., Geneva, Switzerland.
- Rebien, M. (2019). Employers search: Are employee referrals effective? *Economics Bulletin*, 39(4), 2499–2506.
- Redline, C. D. (2011). *Clarifying survey questions*. Doctoral dissertation, University of Maryland, College Park.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly*, 73(2), 325–337.
- Statistics Canada. (2020). Job vacancy and wage survey. Retrieved from <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey%5C&SDDS=5217>
- Tourangeau, R. (1984). Cognitive science and survey methods. In J. T. B. Jabine M.L. Straf & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology. Building a bridge between disciplines* (pp. 73–100). Washington, D.C.: National Academy Press.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge, U.K: Cambridge University Press.
- U.S. Bureau of Labor Statistics. (2020). Job openings and labor turnover survey. Retrieved from <https://www.bls.gov/opub/hom/jlt/home.htm>
- Willimack, D. K., & Nichols, E. (2010). A hybrid response process model for business surveys. *Journal of Official Statistics*, 26(1), 3–24.
- Zuell, C., Menold, N., & Körber, S. (2015). The influence of the answer box size on item nonresponse to open-ended questions in a web survey. *Social Science Computer Review*, 33(1), 115–122.