

Understanding Society

Working Paper Series

No. 2016 – 04

July 2016

**The Impact of Dependent Interviewing Wording
and Survey Factors on Reporting of Change**

Tarek Al Baghal

Institute for Social and Economic Research

University of Essex

Non-technical Summary

Longitudinal surveys allow for studying response change within respondents not possible in cross-sectional studies. Prior studies, however, suggest reports of change in longitudinal studies are potentially error-prone. Proactive dependent interviewing (PDI) reminds respondents of previous answers, asking if there has been any change since the last survey, and is a possible method to reduce errors by assisting recall and reducing cognitive burden. However, the way PDI questions are worded has an impact on reports of change, depending on phrasing as either constancy or change in state. A question asking whether a condition is “still the same” produces far less change reports than when the question asks “has this changed”. Research examining these issues relies only on the recorded survey data. It may be, however, that the change reports and differences of these reports across wording are related to variables not included in the survey data set. In particular, the interaction between the interviewer and respondent may be a driving force in results, but analyses have been limited due to a lack of indicators on this interaction. Research that has coded recorded interviews frequently finds that adding indicators of interviewer and respondent behaviours improve understanding response outcomes. To date, there is a lack of similar insight to how PDI questions influence reports of change, and how differences in PDI wording impact this influence.

This study provides this insight through the behaviour coding of recorded interviews from the third wave of the UK Innovation Panel (IP). The third wave of the IP included an experiment using two PDI versions, asking either “Is this still the same?” or “Has this changed?”. The responses to this question indicate an initial report of change in status or not. For those that report change initially to one of these versions, a follow-up asking what about current status. In a number of instances, these follow-up responses are the same as the previous wave response, indicating a change actually did not occur.

To understand outcomes to the different PDI versions and follow-ups, interviewer and respondent behaviours are coded by the sequence of turns in the question-answer process. Capturing this sequencing provides insight to the extent of interviewer-respondent interaction and where potential problems may occur. Interviewer behaviour codes largely indicate the extent of deviation from the standardised survey script, while respondent behaviours include whether an appropriate answer was given and if any uncertainty was expressed or additional information given. Initial results show that when reporting a change, respondents also provide additional information, generally in explanation. This explanation also provides relevant information to the follow-up question, which may affect the interaction. Further, there is more difficulty in the question-answer process when there are responses to the follow-up that show no actual change, indicated by more deviating behaviours and more interviewer and respondent turns in the sequence. Additional results will be discussed, as well as implications. For example, evidence suggests an avoidance of asking PDI in terms of change.

The Stability of Mode Preferences: Implications for Tailoring in Longitudinal Surveys.

Tarek Al Baghal

Abstract

Proactive dependent interviewing (PDI) is a possible method to reduce errors in reports of change in longitudinal studies, reminding respondents of previous answers while asking if any change occurred. However, little research has been conducted on the impact of PDI question wording. This study examines the impact of PDI wording on change reports, and how these wordings interact with other survey features. Results indicate that asking about change in an unbalanced fashion leads to more reports of change initially than other wordings, but only in a face-to-face survey. Follow-up questions led to final change reports that were similar across all wordings, but this necessitates asking additional questions.

Key words: dependent interviewing, questionnaire design, longitudinal surveys

JEL classifications: C81, C83

Author contact details: talbag@essex.ac.uk

Acknowledgements: This study was funded by Economic and Social Research Council Award no. ES/H029745/1, “Understanding Society and the UK Longitudinal Studies Centre” (Principal Investigator: Nick Buck).

1. Introduction

Panel surveys allow for the studying of phenomena not possible in cross-sectional surveys. In particular, these data measure the amount of change (or lack thereof) at both the macro (aggregate) and micro (individual) levels over time. This change can be continuous change, measured as within and across wave change, or can be the differences between administrations of the survey (wave on wave differences). The unique opportunity to measure change also introduces different data quality issues. The current study examines the effects of differences in dependent interviewing wording and other survey factors on the measurement of wave on wave change.

A variety of studies have shown that measures of change are error prone in panel surveys, with much of the observed change being spurious. In some instances, however, change can be underreported as well (Jäckle 2009; Tourangeau et al. 2000). Such errors are prevalent in both categorical and continuous measures (e.g. Conrad et al. 2009; Hoogendoorn 2004; Jäckle 2009; Lynn and Sala 2006; Young 1989), but some types of questions are more affected than others (Rips et al. 2003; Young 1989). The extent and pattern of errors of change in longitudinal surveys suggest that forgetting or other faults in memory are not sufficient in explaining the problem (Tourangeau et al. 2000). Rather, there are multiple causes for erroneous measures of change, including memory errors, other measurement errors, and/or interviewer and data processing errors (Jäckle 2008).

One proposed method to reduce errors in measuring change in panel surveys is dependent interviewing (Jäckle 2009; Mathiowetz and McGonagle 2000). In dependent interviewing, answers from previous interviews are incorporated into the questionnaire of the current wave to assist with recall, either as part of the questions or as edit checks. However, the best design of dependent interviewing is still an open question, in particular regarding the question wording. This study uses experimental data from a longitudinal study to examine the

impact of wording for one particular form of dependent interviewing. Other survey factors that may also interact with wording, such as mode, question content, and the amount of change experienced are also examined. The goal is to understand the impact of dependent interviewing wording on measuring change and to discuss the implications for improving survey design. Given the possible benefits of dependent interviewing, including the reduction of cognitive burden on respondents as well as the reduction of errors in measures of change, it is important to identify designs that potentially best improve data.

2. Types of Dependent Interviewing

In general, there are two main types of dependent interviewing that have been developed to counter these errors in measuring change: reactive dependent interviewing (RDI) and proactive dependent interviewing (PDI). In RDI, respondents are first asked the survey questions independently and data from the previous waves are used as edit checks. If a substantive change is identified, a follow-up question containing the past information is presented to check whether the change is correct. For example, a respondent who reported having a checking account at a previous wave but not in the current wave may be asked, “Last time we interviewed you, you mentioned having a checking account. Have we missed it?”. RDI may be particularly useful in ensuring consistency of responses and reducing error when used for a limited number of items (Mathiowetz and McGonagle 2000).

PDI by design focuses on improvement and efficiency of the initial recall attempt, and is the focus of the current research. When questions are repeated from wave to wave, asked independently, the respondent must carry out the steps of the response process as they would with any question: comprehension, retrieval, judgment, and response (Tourangeau et al. 2000). By using previous responses in asking the question, PDI shifts the recall task to that of

recognition, and reduces the overall burden to respondents. However, respondents may still not make a complete effort in responding, i.e. satisficing.

PDI reduces the need for recall as it includes the respondent's status at the previous survey in the stem. This information is used in questions as a bounding technique (e.g. Neter and Waksburg 1964) or for asking whether a change has occurred since the last wave (Jäckle 2009, Mathiowetz and McGonagle 2000). Several studies have used PDI for the latter purpose, asking about the respondents' current status relative to previous waves (Conrad et al. 2009, Hoogendoorn 2004, Jäckle 2009, Lugtig and Lensvelt-Mulders 2013, Lynn and Sala 2006, Lynn et al. 2006, Mathiowetz and McGonagle 2000, Rips et al. 2003). These studies suggest that PDI increases data quality by reducing the spurious change frequently found in panel surveys (Hoogendoorn 2004, Jäckle 2009, Lynn et al. 2006, Lynn and Sala 2006).

3. The Cognitive Impact of Proactive Dependent Interviewing

A key aspect that has received little attention in the use of PDI and its effects is how these questions are worded (Mathiowetz and McGonagle 2000). If the question asks about change in different ways, then differences in estimates of change may occur. For example, respondents may be reminded of their previous answer and then asked "is this still the same?" or "has this changed?". The wording may matter because although PDI reduces the cognitive burden for respondent, it also increases the opportunity for them to satisfice (Lugtig and Lensvelt-Mulders 2013; Mathiowetz and McGonagle 2000).

In the case of PDI questions requiring a yes or no response, respondents may satisfice by acquiescing and saying "yes" regardless of whether the implied status is accurate. If respondents satisfice and incorrectly acquiesce in agreement with the PDI question, then the wording could have a significant impact on the measures of change. Specifically, if a PDI question asks if a value is the same or asks if the value has changed, and respondents

acquiesce as suggested, then the former version will indicate significantly less change than the latter. Evidence suggests that this effect may occur, as respondents have been shown to be more likely to endorse the explicit option offered them (Schuman and Presser 1981). This endorsement likely occurs due to the focusing of respondents' attention to a particular outcome (Narayan and Krosnick 1996). The evidence for satisficing in PDI has been mixed. Jäckle (2009) finds that there is not relatively more underreporting in PDI using "still the case" wording, suggesting against satisficing. However, Lugtig and Lensvelt-Mulders (2013) find greater correlated measurement error in PDI (compared to independent measures or RDI), indicating possible satisficing. These studies, however, have not compared the impact of differing wording on potential satisficing or measurement of change.

If the PDI as phrased with yes or no response options increases endorsement of change or no change due to yea-saying, other wording choices may be necessitated. Research on acquiescence has found that when the options of focus are balanced (response options such as "still the same or has this changed" rather than yes/no) acquiescence is minimized (Narayan and Krosnick 1996). The expectation is that yes/no questions worded in terms of constancy will show less change than yes/no questions focusing on change, whereas questions balancing options of constancy and change may likely fall between these two in terms of amount of change.

How PDI wording impacts reporting of change may be influenced by several other factors. Interviewers may have more difficulty administering PDI questions, depending on the design, such as computer programming (Mathiowetz and McGonagle 2000). Further, respondents may be more likely to acquiesce in the presence of interviewers compared to self-administered modes (de Leeuw 2005). If so, then yes responses to PDI questions may be inflated, changing the measurement of change depending on whether the question is phrased in terms of constancy or change. Respondents may be more likely to satisfice generally using

a self-administered mode, such as a web survey, however (de Leeuw 2005; Heerwegh and Loosveldt 2008). Therefore, whether the survey is interviewer- or self-administered is important in understanding the influence of PDI on change measures. Given the possible different directions that interviewer- or self-administration may affect the processing of the question, the comparative impact these modes will have on measures of change is not clear *a priori*.

The content of the question is an additional factor in understanding the potential impact of PDI. Attitudes are more likely to be constructed at the moment the question is asked, than to be recalled as a stored value, as would be the case for autobiographical information (Sudman et al. 1996; Tourangeau et al. 2000). As such, attitudes may be more likely to change over time, and are generally more affected by question wording and survey design than autobiographical questions (Schuman and Presser 1981; Sudman et al. 1996). Given this increased propensity to change and to be influenced by question wordings, attitude questions may be affected by PDI differently. For example, given that attitudes are more prone to change, asking a question explicitly about change rather than constancy may promote change reports. Further, reports of change are more likely questions about some types of events than for others (Tourangeau et al. 2000; Young 1989). These differences occur due in part to differences in retrieval difficulty (Smith and Jobe 1994). Given this increased difficulty, and that PDI is designed in part to reduce retrieval difficulty, it may be additionally expected that questions more prone to change generally will also be differentially influenced by PDI. If responses are prone to change generally, again, asking PDI using change wording may lead to more change for these questions than constancy wording.

Besides PDI design and the survey factors that may interact with the wording, given the cognitive nature of reporting on change, a respondent's cognitive ability may influence the reporting of change independently. People with lower cognitive ability are more likely to

forget change and falsely recognize the previously provided information as being accurate (e.g. Schwarz et al. 1999), and hence, those with lower cognitive ability may be less likely to report change generally. Further, those who have lower cognitive ability may have more difficulty recalling relevant information (Schwarz et al. 1999) or may be more likely to satisfice as the task becomes more difficult (Krosnick 1991). Also, some people have more volatility in their lives, and should legitimately report change more frequently than others. This volatility makes recall cognitively more burdensome (Belli et al. 2013). Life changes and cognitive ability both may therefore be expected to have an additional influence on the reporting of change beyond the effects of PDI and other design factors.

The following analyses focus on the impact of differing question wordings on change, as well as studying if the influence is moderated by the presence of an interviewer, the content of the questions, and whether the question is prone to change reports. Additional factors such as cognitive ability and general volatility in life are analyzed to identify the impact these have on change reports. The findings will provide guidance on best practices in the design of PDI and awareness of how change reports may be impacted by design and respondent characteristics.

4. Data and Methods

The Innovation Panel (IP) longitudinal survey is part of the United Kingdom Longitudinal Household Study (UKHLS). The UKHLS-IP is a vehicle for experimentation regarding aspects of survey design in a longitudinal survey context. It is based on a stratified, multi-stage probability sample of persons and households in England, Scotland, and Wales. At the fourth wave, a refreshment sample was also drawn. At each annual wave, interviews

are attempted with all household members 16 years of age and older.¹ Prior to wave 5, all interviews were conducted by interviewer. At wave 5, a random two-thirds of sample households were allocated to a mixed-mode web and face-to-face design, while the other third were administered the standard single-mode face-to-face design. In the mixed-mode treatment, if any household member did not respond to the web survey within two weeks, an interviewer was sent to attempt a face-to-face interview.

At the initial wave, conducted in 2008, the response rate by original sample members was 51.7%. Interviews at subsequent waves were attempted with households interviewed and those not contacted or classified as “soft” refusals during the immediate previous wave. The wave 3 completion rate amongst wave 1 respondents was 55.9%, producing a net wave 3 response rate of 28.9% (AAPOR RR3). For the original sample, there was a 41.8% completion rate among those who responded at wave 1, for a net wave 3 response rate of 23.9% (AAPOR RR3). For the refreshment sample, the wave 4 response rate (their initial wave) was 48.8% and the wave 5 completion rate amongst wave 4 respondents was 71.7%, producing a net wave 5 response rate of 35.0% (AAPOR RR3).

To examine the impact of dependent interviewing wording, experiments were conducted starting in the third wave using proactive dependent interviewing. Some of the questions were asked in consecutive waves in PDI format. To avoid possible effects that may occur by using PDI responses in later PDI questions, this analysis uses only the first time the respondent was asked a given PDI question. At the third wave, three questions were first asked using PDI: a subjective evaluation of health on a five-point scale, whether the respondent was employed in a permanent or temporary position, and the number of hours usually worked in a week, with both the latter two questions asked only of respondents that had worked in the past week (full question wordings available in Appendix A). Responses

¹ Full details of the sample design and field procedures can be found in the IP user’s guide at <https://www.understandingsociety.ac.uk/documentation/innovation-panel>

from the immediate prior questionnaire were used to fill information in the active wave, and all respondents were reminded of this information through the statement, “The last time we interviewed you on [DATE], you said that” (see Appendix A). Households were allocated to one of two PDI conditions, One half-received immediately after the reminder a question phrased as constancy, “Is that still the same?” (the *unbalanced, still* version), while the other half received a question, “Has that changed?” (the *unbalanced, change* version). Respondents received the same experimental version for all questions.

In Wave 5, the number of questions asked using PDI increased, two new question versions were included, and respondents were assigned randomly to one of the four versions. The new PDI questions included those asking about being an employee or self-employed, sole-owner or joint owner (if self-employed), being paid by salary or by the hour, the mode of transport to work, the type of educational institution currently attending (if a student), and whether the respondent would like to move or not (Appendix A). In addition, the general health rating and hours worked asked using PDI in Wave 3 were also asked using PDI in Wave 5, and are included in the analysis for those who had not responded to these questions in the PDI version previously, i.e. the refreshment sample. A scripting error caused some of the additionally planned PDI questions to not be implemented, including the previously used question on permanency of employment.

Respondents were assigned to one of four PDI wording conditions independent of Wave 3 assignment. Once assigned to a condition, all experimental questions were asked using the same PDI version. In addition to the same constancy and change versions used in Wave 3, two balanced forced-choice questions requiring the respondent to choose constancy or change explicitly are tested. In one version, the question reads “Is this the same or has this changed?” (*balanced, still/change*) whereas the other version reverses the order to “Has this changed or is this the same?” (*balanced, change/still*). In all versions and waves, if the

respondent indicated a change had occurred, a follow-up question is triggered to ask what the new answer is for the question.

In analyzing whether the reporting of change is affected by question wordings, the unit of analysis is all responses to PDI questions within respondents. Two measures of change are indicated by the data. First, initial change measures are obtained based on the response to the assigned PDI versions. Second, final change measure are based on the final reports of the respondents when asked the follow-up question. If a respondent indicated a change to the PDI question, they were asked the follow-up question for the new status value. Some respondents provided the same value as the preloaded information, that is, no change actually occurred based on the final report. If a respondent had indicated change in PDI, but gave the same response in the follow-up, their initial report would be coded as a change and the final report coded as no-change. For the hours worked question, the only continuous-type measure, a ten-percent cutoff is used, consistent with past research on dependent interviewing (Jäckle 2009; Lugtig and Lensvelt-Mulders 2013). If a respondent said their hours worked had changed to the PDI question, but provided a value that changed less than 10% of the previous value, they would be coded as no-change in the final change measure.

The PDI questions about self-assessed health and the desire to move are attitudinal in nature, whereas the remainder asked about autobiographical information. The PDI question about the number of hours worked per week is the only continuous type of measure available; the remainder is all categorical in nature. An indicator for whether a change had occurred for the particular questions used in PDI by comparing answers of the immediate two previous waves, where the questions had been asked independently. If there was a difference, a previous change was indicated, while no difference indicated no prior change.

Besides the PDI question wordings and the survey factors that may interact with these, several other variables may be of interest. The sum of several possible life changes

between waves was calculated as a measure of volatility experienced by a respondent. These include whether the number of times a respondent moved, changed marital or relationship status, had a child, added any educational qualifications, or suffered or overcame a long-term illness or disability, all within the past year of the interview. Additionally, gender and race (white/non-white) are included, as are age at the time of the survey and education, which may be proxies for cognitive ability (Schwarz et al. 1999). An additional measure for cognitive ability comes from the subjective rating of the interviewer if the respondent understood the questionnaire or not, completed after the interview. For those responding via the web survey, the interviewer measure for understanding from the previous wave is used. Based on the design of the web, mixed-mode survey, to account for possible selection effects, significant correlates to the mixed-mode design identified in Jäckle et al. (2013) are included as respondent characteristics. The only variables that were jointly significant in predicting individual response rates are urban location and saying there was definitely no chance of responding to a web survey (recorded in Wave 4).

5. Results

The composition of the IP sample in regards to the respondent characteristics of interest is presented in Table 1. Females are overrepresented, and a somewhat older population is represented. The sample is also largely white, living in urban areas, and indicated by the interviewer as understanding the questionnaire. A sizable minority indicated they would be unlikely to respond via the internet. Respondents had on average less than one life change in the past year, but there was a large range. The majority of respondents (69.91%) had zero life changes in the year prior to the given survey wave, another 24.95% had one life change, while 5.14% had between two and five.

Table 1. Respondent Characteristics

Variable	Mean/Proportion of Sample
Age	47.83
Female	0.56
College/Professional Degree	0.29
White	0.85
Urban	0.75
Understand Questionnaire	0.96
Unlikely to Respond by Web	0.29
Number of Life Changes, Past Year	0.37

It is assumed that PDI will reduce the amount of change observed relative to independent asking of the questions. To test if this assumption holds true in the IP data, the immediate two waves previous to the PDI question being asked where the question was asked using independent interviewing was used to calculate change. Change is measured from the first and second waves for Wave 3 PDI questions and third and fourth waves for Wave 5 PDI questions. Change was identified if the responses between the two independent reports do not match (using the 10% threshold for the hours worked question noted above). The change in all items across independent asking of the questions is 22.62%; conversely, the initial responses to the PDI questions indicate an 11.28% change in responses. Controlling for the repeated measures within-individuals, the difference is statistically significant, $F(1,1223) = 277.81, p < 0.001$. Since PDI is used to reduce possible spurious change, this reduction is appropriate and suggests that the PDI is performing according to design.

Given that PDI does reduce change generally, as expected, the question remains as to whether wording of the PDI impacts change outcomes, and if the impact differs based on survey features. There are 7,152 responses obtained from 1,867 respondents using the PDI questions (range of number of responses given: 1-7). While 11.28% ($n=807$) of the initial responses to PDI indicated change, when examining change based on the final outcome,

7.59% (n=543) responses indicated a change across waves. The remaining 3.69% (n=264) of initial reports of change are responses initially indicating a change in the PDI question, but were followed-up with responses the same as the previous wave, i.e. indicating no change in the final report. These discordant responses do not appear to be due to a small number of respondents making this error multiple times; 235 respondents accounted for the 264 responses, with a median of 1 and a slightly higher mean of 1.13. Table 2 breaks down the percentage of initial and final reports of change for the differing wordings across mode (face-to-face and web), question content (attitude and factual), whether there was a change reported at between the immediately two previous waves (denoted “Previous Change”).

Table 2. Percentage of Reported Change, Initial and Final Response

	<i>Unbalanced, Still</i>		<i>Unbalanced, Change</i>		<i>Balanced, Still/Change</i>		<i>Balanced, Change/Still</i>	
	initial	final	initial	final	initial	final	initial	final
<i>Mode</i>								
Face-to-Face	9.1	7.2	19.9	9.3	9.3	8.5	10.0	8.1
Web	6.4	6.2	6.8	3.8	5.3	4.9	8.8	8.0
<i>Question Content</i>								
Attitude	8.8	7.4	17.2	9.3	9.2	8.6	11.8	9.4
Factual	8.4	6.6	17.2	7.2	7.0	6.3	7.8	7.0
<i>Previous Change</i>								
No Previous Change	7.0	5.6	15.4	6.2	7.1	6.4	9.4	7.8
Previous Change	15.6	13.3	24.6	16.6	22.9	22.9	12.5	12.5
Total	8.6	7.0	17.2	8.2	8.0	7.3	9.6	8.1
n	2353		2198		1334		1267	

The data show that, in general, there does not appear to be much difference in reports of change, either initial or final, between attitude and factual measures across all PDI versions, with no significant differences identified. However, there are significant differences if a change occurred in the previous wave or not and by the mode of the survey. When a change was indicated in the previous wave, there was a higher reported change across all PDI wordings, $F(1,632)=74.34$ $p<0.001$. Similarly, generally reports of change are greatest in face-to-face survey and lower in web surveys $F(1,255)=45.94$ $p<0.001$. These differences are reduced somewhat when examining final reports of change, but still significant. The web version also produces lower levels of change within every version of the PDI question, for both initial and final reports, although not as large of differences in the final reports.

Importantly, there are differences between the PDI wordings. The *balanced, change/still* version appears to increase change reports slightly over the *balanced, still/change* version, adding possible evidence of possible satisficing (via a primacy effect). However, the overall total difference between balanced versions is not significant, $F(1,854) = 2.09$ $p=0.15$. The PDI *unbalanced, change* wording, however, leads to substantially greater amounts of initial reports of change than any of the other versions. This difference holds across measurement types and content; however, this pattern does not arise for both modes. The *unbalanced, change* wording produces comparable amounts of change as the other PDI versions when asked by web, $F(3,546) = 1.56$ $p=0.20$. Rather, the effect of the *unbalanced, change* wording instead comes from differences in the face-to-face survey. That the differences between the *unbalanced, change* and other versions occur only in the face-to-face mode, with the most initial change reported while there is no difference in final change reports is consistent with two possible mechanisms.

First is the possibility that respondents are satisficing by acquiescing in the presence of the interviewer. This would explain why initial reports of change are only higher than

other wordings in the face-to-face survey. A second possibility is that interviewers and/or respondents have difficulty with the *unbalanced, change* version of PDI: the negative response (“no”) indicates that the current status is the same as it was in the previous wave. Further, a few living status questions early in the survey are phrased as to whether a household member *still* has the same living situation, where a positive response indicates the status is unchanged. If interviewers or respondents have difficulty having to answer in the negative to indicate the same status, possibly due to previous experience in the survey instrument, initial change reports would be inflated.

Examining only responses in the *unbalanced, change* wording version conducted in the face-to-face survey shows that 198 responses were corrected from the initial change report in PDI to no change in the follow up. These corrections in the follow up represent 75.0% of the total number of such corrections (n=264) across all versions and modes. To examine whether the amount of change reports are related to respondents or interviewers, intraclass correlations (ICC) are calculated for interviewers and respondents (accounting for clustering within interviewers). The ICC in regards to initial change reports in the *unbalanced, change* version only is 0.068 for interviewers and for respondents, 0.159. These ICC suggest that while both interviewers and respondents may contribute to the greater change reports in this version, more effect is attributable to respondents. While it is not clear given the data what the exact mechanisms leading to higher reports of change are, it does not appear to be mainly due to interviewer error, although this error may contribute. Rather it appears more due to respondent effects, suggesting the greater possibility of respondent acquiescence in this instance.

To examine the possible interaction between PDI wording and survey features, multivariate analyses are conducted. Models are estimated for both initial and final change report as outcome measures. Given that the dependent variables are measured at the response

level, always nested within respondents (second-level), multilevel logistic regression models are estimated, with the outcome being 1 if a change is recorded, 0 if there is no change.

As a first step, random-intercepts only (i.e. null) models are estimated to calculate variance components and the intra-class correlation (ICC) coefficients, restricted to cases used in the full models to allow for better comparison. The respondent ICC for the initial change null model is 0.088 and for the final change null model it is 0.071. These ICC indicate that while respondents contribute to change reports, much of the variation is explained by factors at the question-level. Models including both question and respondent characteristics are estimated as a next step to identify the factors that influence change reports. Question characteristics include the differing question wording versions, using the *unbalanced, still* as the reference category, the mode, if the question is attitudinal or factual, and whether there was a reported change for the question at the previous wave, measured as a t-1 lag of change for each of the individual questions. Interactions between question wordings and these features are also included to test whether the survey features impact the effects of the PDI wording.

Respondent characteristics are those presented in Table 1: age, sex, race (white or not), education (college/professional degree or not), living in an urban area or not, whether the respondent said they would not respond via the internet, and the interviewer subjective evaluation of whether the respondent understood the questionnaire. Given that respondents with more change in their life history may be more expected to change and have more cognitive difficulty accurately recalling change (Belli et al. 2013), the number of changes in a respondent's life using the measure discussed above is also included

For both initial and final change the full model, including respondent characteristics, improves model fit over models including only question characteristics: for the initial change model, $\chi^2_8 = 25.96$, $p < 0.05$, and for the final change model $\chi^2_8 = 34.793$, $p < 0.05$. Table 3

presents the full model estimates for initial and final change. In order to better illustrate the impact interactions have, predicted probabilities from each model are calculated, taking into account the random effects. Figures 1 and 2 display the mean predicted probabilities of both initial and final change reports change reports estimated from the models in Table 3. Figure 1 displays predicted probabilities by wording and survey mode. Figure 2 displays the predicted probabilities by the question wording and whether a change occurred in the previous wave for a given question.

Table 3. Multilevel Estimates for Initial and Final Change Reports

	Initial Change	Final Change
Question Characteristics		
<i>PDI Wording (Unbalanced, Still)</i>		
Unbalanced, Change	1.111*	0.357
Balanced, Still/Change	0.070	0.300
Balanced, Change/Still	0.187	0.389
Web Response	-0.157	0.098
Attitude Question	-0.129	0.060
Lagged Change	0.905*	0.940*
<i>Interaction w/Web Response</i>		
Unbalanced, Change	-1.015*	-0.924*
Balanced, Still/Change	-0.329	-0.560
Balanced, Change/Still	0.015	-0.119
<i>Interaction w/Attitude Questions</i>		
Unbalanced, Change	0.023	0.028
Balanced, Still/Change	0.279	0.155
Balanced, Change/Still	0.594*	0.275
<i>Interaction w/ Lagged Change</i>		
Unbalanced, Change	-0.557*	0.014
Balanced, Still/Change	0.186	0.279
Balanced, Change/Still	-0.696	-0.461
Respondent Characteristics		
Age	-0.005	-0.008*
Female	0.026	0.154
White	-0.160	-0.091
College/Professional Degree	-0.040	-0.015
Unlikely to Respond by Web	-0.024	-0.017
Urban	0.012	-0.081
Life Changes	0.238*	0.309*
Understand Questionnaire	-0.204	-0.130
Constant	-2.153*	-2.634*
<i>Random-effects Parameters</i>		
Respondent Variance	0.222	0.168
ICC	0.063	0.049

*p<0.05 Responses = 6518 Respondents = 1620

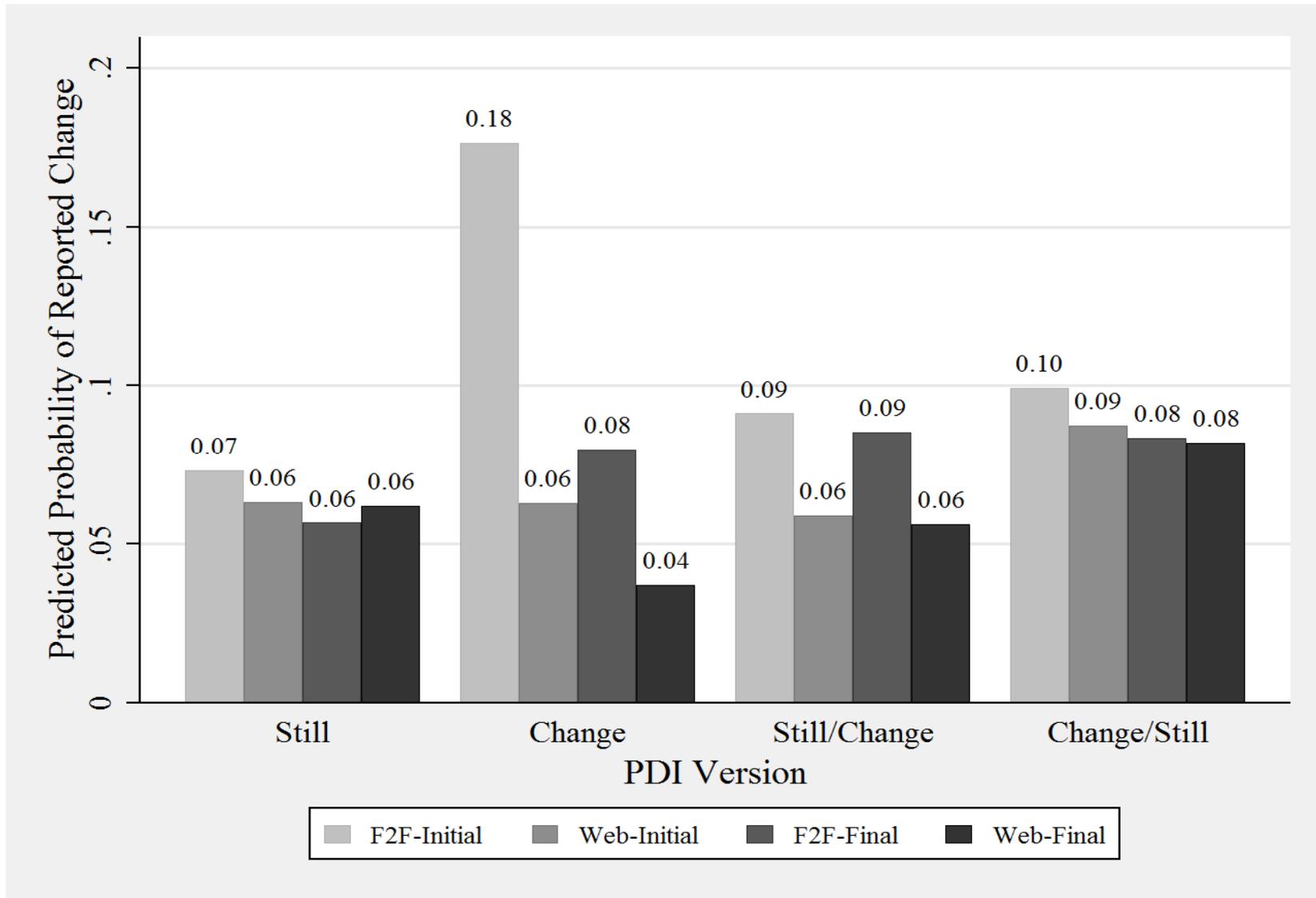


Figure 1. Predicted Probabilities of Initial and Final Change Reports, by PDI Version and Mode

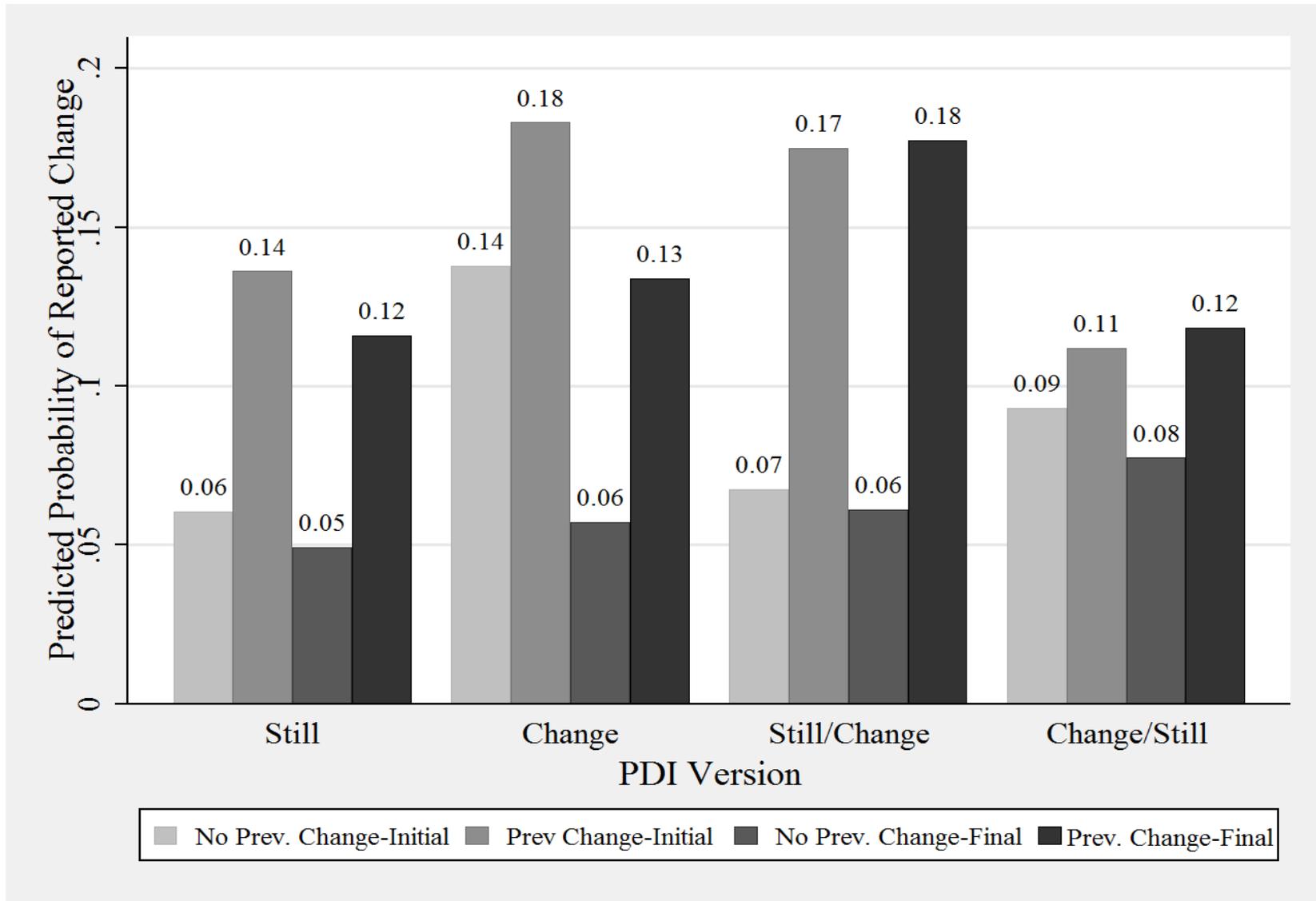


Figure 2. Predicted Probabilities of Initial and Final Change Reports, by PDI Version and Previous Change Report or Not

These results shed light on the impact the PDI wordings and survey factors have on reporting change in response to PDI (i.e. initial change report) and on the final change report. In regards to the report of change to the initial asking of the question, taking into account all effects in the full model suggest that the *unbalanced, change* version increases the probability of an initial change report in the face-to-face mode, regardless of other question characteristics. Conversely, there is almost no effect of the *unbalanced, change* version in the web survey, with effects similar to the other question versions for all question characteristics. These differences are illustrated in Figure 1, where the mean predicted probability for the *unbalanced, change* version in the face-to-face mode is 0.18, and is much higher than any other version in either mode. Conversely, the predicted probabilities of initial reported change in the web survey are similar across all wordings (*unbalanced, still* = 0.06; *unbalanced, change* = 0.06; *balanced, still/change* = 0.06; *balanced, change/still* = 0.09). This overall increase only when the interviewer is present is consistent with the hypothesis that the *unbalanced, change* version potentially increases acquiescence, inflating the estimate of overall change.

The impact of the *unbalanced, change* version is moderated by other question characteristics. Question content (attitude or factual) has no significant impact. However, a change in the previous wave significantly increases the estimated probability of change for all question wordings, and this is also the case for the *unbalanced, change* version. Combining all the main and interactive effects for the *unbalanced, change* wording is used in the face-to-face survey and there is a prior change leads to an odds ratio of $e^{0.348}=1.42$ for an initial change report. Although still a positive increase, the impact of a previous change is somewhat less in the *unbalanced, change* version than in either the *unbalanced, still* or *balanced, still/change* versions, as indicated in Table 3 by the negative effect found for the interaction between the *unbalanced, change* version and lagged change.

These effects are illustrated in Figure 2, which includes the predicted probabilities of initial change reports for the PDI wording when there was a previous change or not. The predicted probability of an initial change report under the *unbalanced, change* version when there is a previous change is 0.18, significantly higher than 0.14 estimated probability for the same wording but when no change had occurred previously. However, as seen in Figure 2, this difference in predicted probabilities is less than for the *unbalanced, still* or *balanced, still/change* versions, indicating the interactive effect previous change has with the *unbalanced, change* version. In all instances, when there was a change indicated at the previous wave the estimated probabilities are of a change report are greater than when no change was reported. In every case, the predicted probability of a change is greater than 0.10; in comparison, only two estimated probabilities are 0.10 or more when examining mode effects in Figure 1.

The only other significant effect involves the *balanced, change/still* question wording. While this wording does not significantly affect the estimated probability differentially across modes or when there is a previous change indicated, there are differences in the impact across question content. Impacts of other PDI wordings are not influenced by whether the question is attitudinal or factual, but the *balanced, change/still* version leads to higher estimated probabilities of change when the question asks about an attitude compared to when it asks about factual information. The limited nature of the impacts of this wording and of question content only interacting with this wording does not allow for definitive conclusions to be made. However, this question wording leads with the “change” option, and the increase in estimated probability of initial reports of change could point toward possible primacy, where this impact is accentuated only when asking about an attitude. If primacy is the cause, then it would be expected to have lower change in the still version(s); however, while directionally this is the case, the difference for the *balanced, still/change* version is not significant.

The full initial change model has a respondent ICC of 0.063, a 28.4% relative decrease compared to the null model. Most of the respondent characteristics do not have a statistically significant impact on the probability of reporting change initially. Only one is significant: the number of life changes in the past year, suggesting more changes in other parts of life increases the chance of an initial change report. It may be that those with more changes in life overall will also have a greater chance of change for any given question, or may be related to the fact that the number of changes in life may increase the difficulty of recalling transitions (Belli et al. 2013). Measures of cognitive ability (age, education, and understanding of the questionnaire) have no significant impact on estimated probability, contrary to expectation.² Given that the probability of reporting an initial change increases with changes in life and reports of change for the same in the past for a given question suggests the effect of life changes indicates a respondent's tendency towards change generally rather than cognitive difficulty in recall.

In examining respondents' final change, the incorporation of all of the variables in the full model for final change in Table 3 has a decrease in the estimated ICC from the null model of 31.0%, similar to the reduction in the initial change model. Unlike the initial change model, however, there are few response-level effects. PDI wording does not impact final change reports in the same way as initial change reports. The one exception is in a web survey, asking the *unbalanced, change* version reduces the estimated likelihood of change relative to other wordings. This relative reduction for the *unbalanced, change* version in the web survey holds regardless of question content or whether a change occurred in the previous wave or not. This decrease is seen in Figure 1, where the predicted probability of a final change report for the *unbalanced, change* version in the web survey is 0.04.

² Measuring age in categories instead of continuously also shows that older respondents are not significantly different from younger as a category. As such, the continuous measure is maintained.

The direction and significance for the interaction between the *unbalanced, change* wording and mode is the same as what was found in the initial change model, although the overall effect is different given the significant positive effect of the *unbalanced, change* wording found in the initial change model. Also similar to the initial change model, if a change was reported in the previous wave there is a significant increase the estimated probability of a final change report, for all wordings, across modes and contents, which can be seen in Figure 2. Like the initial change model, those with greater life changes have a higher probability in reporting a final change. These effects are consistent with the argument that changes in status generally suggest a respondent is more likely to report a change across waves.

6. Discussion and Conclusions

The collection of survey data from the same individuals over time offers opportunities to researchers interested in a variety of issues that could not be studied in a cross-sectional survey. Measures of change for a number of phenomena can be collected at both the macro- and micro-levels; however, spurious change is frequently found across waves of the survey. These errors in change measures are due to a number of factors, including recall errors, other measurement errors, and/or interviewer and data processing errors. One design option suggested to reduce such errors is dependent interviewing, of which several forms exist. Little research has examined the design features of dependent interviewing and how these may interact with other features of the survey and questions being asked. The current research adds to the extant literature by examining the effect of wording for one form of dependent interviewing, proactive dependent interviewing (PDI) and the wording's relationship with other survey features and respondent characteristics on reports of change.

As with many other surveys, the change in responses found here across waves is significantly greater when asking the questions independently (i.e. simply repeating the same question at different time points) than when using a dependent interviewing approach. The question then arises as to whether wording of PDI influences reports of change. Importantly, asking the PDI as to whether a response has changed or not in an unbalanced format in an interviewer-administered survey leads to more reports of initial change than an unbalanced version asking whether the response is still the same or balanced versions asking whether the answer is still the same or changed (or with the options reversed). This increase in change reports holds across different question content and whether a change was reported for that question in the past or not. That is, the *unbalanced, change* wording version of PDI in the face-to-face survey appears to reduce spurious change less than other PDI versions.

The *balanced, change/still* version also leads to a greater probability of an initial change report when asking about attitudes compared to factual information. Why this wording only impacts attitudes is not wholly clear; generally, there are no differences in change reports for attitudes or factual reports across the PDI versions, except this one. The fact that it occurs with attitudes leading to more change is consistent with expectation, given the more temporal nature of attitudes and the fact that attitudes questions are more sensitive to wording. Taken together, it appears that asking about change as the lead response option may induce unexpected and unwanted reports of change on the initial asking of the question, and should be avoided.

However, it is useful to note that the final data appears less impacted by question wording than does the initial report. There are two important caveats to this however, which further guides wording choice. First, is the exception found in the final reported change model where *unbalanced, change* interacts with the web mode, leading to reductions in change reports. Again, the outlier is the *unbalanced, change* version. Second, even if in the

end most responses were corrected to a level consistent with other wordings, this always required the additional follow-up question. Using PDI questions that lead to higher reports of initial change in the least increases the number of questions asked in a survey, potentially adding to respondent and interviewer fatigue and cost. Along with the results for the initial change model, the implication is that survey design should avoid PDI wording asking about change as the lead option.

These results also shed some light on the possible cause of these inconsistencies between initial and final data. The reductions in initial to final change are due to the number of respondents that indicated the same response as they had in the previous wave even though they had just indicated that a change had occurred to the initial question. These discordant responses occurred most frequently in face-to-face survey with the *unbalanced, change* wording of the PDI, highlighting the problem in particular with this wording. Although these inconsistencies occur mostly in the interviewer-administered version of the survey, the intraclass correlations suggest that the effect is more clustered within respondents, rather than interviewer, although both contribute to the observed variance. It appears that the impact of this wording is due to interviewer-respondent interaction, possibly the respondent acquiescing in the interviewer's presence, answering "yes" as a default, leading to greater than expected change in the *unbalanced, change* version in a face-to-face survey. It may also be that the wording impact is related to the survey design, where prior unbalanced questions asking about if the status was the still the same may have caused confusion.

Finally, the current research adds to a growing literature showing the impact that item- or response-level characteristics have on outcomes in conjunction with respondent effects, suggesting the need for a multilevel approach in survey analysis (e.g. Yan and Tourangeau 2008; Couper and Kreuter 2013; Al Baghal et al. 2014). The importance of response-level characteristics is highlighted by the significant effects found at this level,

whereas far fewer respondent characteristics are significant. This is not to say that respondent effects are not important; respondents still account for some of the observed variance in change reports. Age is borderline significant in the initial change model, and is significant for final change; both results suggests younger respondents report more change. Importantly, the number of life changes in the past year has a significant relationship with reports of both initial and final change as do changes reported for the given question in the past wave. Taken together, these effects are evidence that those with more volatility in their lives may be expected to change more on any given response across waves. These respondents may have more true change to report, or possibly that those who have more change generally have difficulty accurately reporting change, leading to greater misreporting.

While this research has highlighted the importance of PDI wording and survey design on the reports of change, some limitations to the study should be noted. First, the survey design used a PDI variant asking about whether the status for some household living questions was the same for all previously interviewed respondents earlier in the interview. By doing so, respondents and interviewers may have become accustomed to the *unbalanced, still* wording. While the relatively few questions asked using this constancy wording and the spacing between questions suggests this may be unlikely, it cannot be ruled out. Second, there are other potential design features that may be of importance that are not possible to test given these data. For example, continuous measures may be more cognitively demanding than categorical measures (e.g. Bradburn and Miles 1979), and thus may interact with PDI differently. However, with only one continuous measure it was not possible to disentangle effects. Further, while age, education, and an interviewer observation are included to examine cognitive ability, these are only proxies, and potentially better measures may have indicated other effects.

Appendix A: Questions Used

Attitude Questions

- In general, would you say your health is ...
 - Excellent
 - Very Good
 - Good
 - Fair
 - Poor

- If you could choose, would you stay here in your present home or would you prefer to move somewhere else?
 - Stay Here
 - Prefer to Move

Factual Questions

- How is your pay calculated, in particular are you salaried or paid by the hour?
 - Salaried
 - Basic salary plus commission
 - Paid by the hour
 - OtherAre you an employee or self-employed?
 - Employee
 - Self-employed

- Are you...
 - At School
 - At Sixth Form College
 - At Further Education (FE) College
 - At Higher Education (HE) College
 - or at University?

- Are you working on your own account or are you in partnership with someone else?
 - Own account (sole owner)
 - In partnership

- And how do you usually get to your place of work?
 - Drive myself by car or van
 - Get a lift with someone from household
 - Get a lift with someone outside the household
 - Motorcycle/moped/scooter
 - Taxi/minicab
 - Bus/coach
 - Train
 - Underground/Metro/Tram/Light railway
 - Cycle
 - Walk
 - Other

- Leaving aside your own personal intentions and circumstances, is your job...
 - A permanent job
 - Or is there some way that it is not permanent?

- Thinking about your (main) job, how many hours, excluding overtime and meal breaks, are you expected to work in a normal week?

Proactive Dependent Interviewing (PDI) Question Wording Structure

Unbalanced, Still

The last time we interviewed you on [DATE] , you said that ... was [PREVIOUS RESPONSE] . Is that still the case?

- Yes
- No

Unbalanced, Change

The last time we interviewed you on [DATE] , you said that ... was [PREVIOUS RESPONSE] . Has that changed?

- Yes
- No

Balanced, Still/Change

The last time we interviewed you on [DATE] , you said that ... was [PREVIOUS RESPONSE] . Is that still the case or has that changed?

- Still the case
- Has changed

Balanced, Change/Still

The last time we interviewed you on [DATE] , you said that ... was [PREVIOUS RESPONSE] . Has that changed or is that still the case?

- Has changed
- Still the case

References

- Al Baghal, T., R.F. Belli, A.L. Phillips, and N. Ruther. 2014. "What Are you Doing Now? Activity Level Responses and Error in the American Time Use Survey." Forthcoming at *Journal of Survey Statistics and Methodology*
- Belli, R.F., I. Bilgen, and M.T. Al Baghal. 2013. "Memory, Communication, and Data Quality in Calendar Interviews" *Public Opinion Quarterly* 77: 194-219
- Conrad, F.C., L.J. Rips, and S.S. Fricker. 2009. "Seam effects in quantitative responses." *Journal of Official Statistics* 25: 339-61

Couper, M. P., and F. Kreuter. 2013). "Using Paradata to Explore Item Level Response Times in Surveys." *Journal of the Royal Statistical Society: Series A*, 176: 271-286.

de Leeuw, E.D. 2005. "To mix or not to mix data collection modes in surveys." *Journal of Official Statistics*, 21: 233-255.

Heerwegh, D. and G. Loosveldt. 2008. "Face-to-face versus web surveying in a high-internet-coverage population." *Public Opinion Quarterly* 72: 836-846

Hoogendoorn, A.W. 2004. "A questionnaire design for dependent interviewing that addresses the problem of cognitive satisficing." *Journal of Official Statistics* 20: 219-32

Jäckle, A. 2008. "Dependent interviewing: Effects on respondent burden and efficiency in data collection." *Journal of Official Statistics* 24: 411-30.

Jäckle, A. 2009. "Dependent interviewing: A framework and application to current research." In *Methodology of Longitudinal Surveys*, edited by P. Lynn, 93-112. Chichester, England: John Wiley.

Jäckle, A., P. Lynn, and J. Burton. 2013. "Going Online with a Face-to-Face Household Panel: Initial Results from an Experiment on the *Understanding Society* Innovation Panel." *Understanding Society Working Paper Series No. 2013 – 03*. Institute for Social and Economic Research, University of Essex

Krosnick, J. A. 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied Cognitive Psychology* 5: 213-36.

Lugtig, P. and G.J.L.M Lensvelt-Mulders. 2014. "Evaluating the Effect of Dependent Interviewing on the Quality of Measures of Change." *Field Methods* 26: 172-190

P. Lynn and E. Sala. 2006. "Measuring change in employment characteristics: the effects of dependent interviewing." *International Journal of Public Opinion Research* 18: 500-509

Lynn, P., A. Jäckle, S.P. Jenkins, and E. Sala. 2006. "The effects of dependent interviewing on responses to questions on income sources." *Journal of Official Statistics* 22: 357-84.

- Mathiowetz, N. A. and K.A. McGonagle. 2000. "An assessment of the current state of dependent interviewing in household surveys." *Journal of Official Statistics* 16: 401–18
- Narayan, S. and J. Kronsick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60: 58-89.
- Neter, J., and J. Waksberg. 1964. "A Study of Response Errors in Expenditures Data from Household Interviews." *Journal of the American Statistical Association*, 59: 18-55.
- Rips, L. J., F.G. Conrad, and S.S. Fricker. 2003. "Unraveling the seam effect in panel surveys." *Public Opinion Quarterly* 67: 522–54
- Schuman, H. and S. Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA.: Sage.
- Schwarz, N., D. Park, B. Knäuper, and S. Sudman. (1999). *Aging, Cognition, and Self-reports*. Washington, D.C.: Psychology Press.
- Smith, A. F., and J.B. Jobe. 1994. "Validity of reports of long-term dietary memories: Data and a model." In *Autobiographical memory and the validity of retrospective reports*, edited by N. Schwarz and S. Sudman. Berlin: Springer-Verlag.
- Sudman, S., N.M. Bradburn, and N. Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.
- Tourangeau, R., L.J. Rips, and K. Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Yan, T., and R. Tourangeau. 2008. "Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times." *Applied Cognitive Psychology*, 22: 51-68.
- Young, N. (1989). "Wave-Seam Effects in the SIPP." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*. 393-398.