2016/17 BACCALAUREATE AND BEYOND (B&B:16/17) MAIN STUDY

OMB # 1850-0926 v.3

Appendix C
Results of the B&B:16/17 Field Test Experiments

Submitted by
National Center for Education Statistics
U.S. Department of Education

April 2017

Two experiments were included in the 2016/17 Baccalaureate and Beyond Longitudinal Study Field Test (B&B:16/17 FT). Full details of the experiments were described and approved in B&B:16/17 Field Test (OMB# 1850-0926 v. 1) Supporting Statement Part B. The first experiment focused on response rates, reducing nonresponse bias, and minimizing sampling design weight variation (see section C.1). The second experiment focused on minimizing measurement error to further improve data accuracy (see section C.2).

C.1 Evaluation of Experiment #1: Finding the optimum strategy to minimize sampling design weight variation and nonresponse bias and boost response rates

The sampling approach for B&B:08/09 Full Scale (see https://nces.ed.gov/pubs2014/2014041.pdf) included a subsample of approximately 10 percent of the National Postsecondary Student Aid Study (NPSAS) base-year interview nonrespondents among the potential B&B-eligible cases. For B&B:16/17, we wanted to explore the feasibility of increasing the subsampling rate in order to minimize sampling weight variation. Therefore, in the B&B:16/17 FT we wanted to test whether we can increase response rates and minimize nonresponse bias, but not at the expense of increased nonresponse variance. The B&B:16/17 FT was comprised of all base-year interview respondents and all base-year interview nonrespondents. We separated the sample into four groups targeted for different intensities of data collection protocols: two groups of base-year interview nonrespondents and two groups of base-year interview respondents. At the same time, we conducted an observational Mahalanobis distance modeling procedure that helped us better understand how individual distances change over time, as data collection proceeds. Mahalanobis distances measure the multivariate distance between the baseline respondent average and an individual nonrespondent and help identify cases most likely to contribute to nonresponse bias. See https://en.wikipedia.org/wiki/Mahalanobis_distance for more technical information.

a. Field Test Design and Response Rates

The sample was split into four groups, which included two groups of base-year interview nonrespondents, randomly assigned to either an aggressive or to a default data collection protocol, and two groups of base-year interview respondents, divided into early and late base-year interview respondents. Base-year interview late respondents received the default data collection protocol, while base-year interview early respondents received a relaxed data collection protocol. All groups received an initial email and letter and reminder emails and postcards to complete the survey throughout data collection, but the offer of a prepaid incentive, outbound calling prompting efforts, and the abbreviated interview differed according to the assigned data collection protocol (aggressive, default, or relaxed).

- NPSAS:16 FT nonrespondents: Aggressive protocol (Group 1)
- NPSAS:16 FT nonrespondents: Default protocol (Group 2)
- NPSAS:16 FT respondents: Late respondents–default protocol (Group 3)
- NPSAS:16 FT respondents: Early respondents–relaxed protocol (Group 4).

Table C.1 depicts the design implemented in the B&B:16/17 field test.

Table C.1. B&B:16/17 field test data collection protocols by data collection phase

Phase of B&B:16/17 FT	NPSAS:16 FT nonrespondents		NPSAS:16 FT respondents	
	Aggressive protocol (Group 1)	Default protocol (Group 2)	Late respondents--default protocol (Group 3)	Early respondents--relaxed protocol (Group 4)
Early completion (4 weeks)	• $10 prepay incentive • Initial letter and email • Begin outbound calling after 2 weeks	• No prepay incentive • Initial email and letter • Email reminders	• No prepay incentive • Initial email and letter noting past participation • Email reminders	• No prepay incentive • Initial email and letter noting past participation • Email reminders
Production (10 weeks)	• Frequent email reminders • Postcard reminders • Abbreviated interview	• Begin outbound calling • Frequent email reminders	• Begin "light CATI"* outbound calling • Frequent email reminders	• No outbound calling • Frequent email reminders
Nonresponse conversion (4 weeks)	• Frequent email reminders • Postcard reminders	• Frequent email reminders • Postcard reminders • Abbreviated interview	• Frequent email reminders • Postcard reminders • Abbreviated interview	• Frequent email reminders • Postcard reminders
Incentive amount	$10 prepay incentive + $20 upon interview completion	$30 incentive upon interview completion	$30 incentive upon interview completion	$20 incentive upon interview completion

* Outbound calling is considered "light CATI" when a minimal number of phone calls placed to sample members is intended mainly to prompt the web response rather than obtain a telephone interview. During the B&B:16/17 FT, these individuals only received approximately half as the calls compared to the default CATI protocols.

The B&B:16/17 FT data collection results provide insight in preparation for the full-scale study regarding the effectiveness of the various protocols in terms of rates of survey completion, nonresponse bias and differences in substantive responses.

Results. Table C.2 provides an overview of the response rates and the associated test statistics for each data collection protocol by data collection phase among NPSAS:16 FT nonrespondents. At the end of data collection, the base-year interview nonrespondents with the aggressive protocol (Group 1) had a significantly higher response rate (37%) than the base-year interview nonrespondents (Group 2) with the default protocol (25%) ($t(2,097) = 3.52$, $p < 0.001$).

Table C.2: T-Test of response rates per phase by experimental condition (in percent)

	NPSAS:16 FT nonrespondents			
Phase of B&B:16/17 FT	**Aggressive protocol (Group 1) response rate (percent)**	**Default protocol (Group 2) response rate (percent)**	**t-value**	**p-value**
Early completion	8.4	4.4	2.29*	0.0223
Production Phase	22.7	12.1	3.67***	0.0002
Nonresponse Conversion Phase	9.6	9.7	-0.02	0.9844
Total	37.2	25.0	3.52***	0.0004

Note: * p<0.05, ** p<0.01, *** p<0.001. Response rates might not add up to total due to varying case base per phase.

Table C.3 presents the cumulative response rates per and the associated test statistics at the end of each data collection phase.

Table C.**3: T-Test of cumulative response rates per phase by experimental condition (in percent)**

	NPSAS:16 FT nonrespondents			
Phase of B&B:16/17 FT	**Aggressive protocol (Group 1) response rate (percent)**	**Default protocol (Group 2) response rate (percent)**	**t-value**	**p-value**
Early completion	8.4	4.4	2.29*	0.0223
Production Phase	29.6	16.2	4.37***	0.0000
Total	37.2	25.0	3.52***	0.0004

Note: * p<0.05, ** p<0.01, *** p<0.001. Response rates might not add up to total due to varying case base per phase.

NPSAS:16 FT interview nonrespondents were randomly assigned to different data collection protocols—that is, Group 1 received the aggressive protocol and Group 2 the default protocol—allowing for an assessment of the effects of different interventions in those protocols on response rates. In the early completion phase (phase 1), the survey protocol for Group 1 (NPSAS:16 FT interview nonrespondents—aggressive protocol) differed from that for Group 2 (NPSAS:16 FT interview nonrespondents—default protocol). Group 1 received a prepaid incentive and was contacted via telephone. In order to investigate whether the prepaid incentive and the addition of telephone as a survey mode increased response rates, we compared the results at the end of the first data collection phase for Group 1 to those of Group 2, who received a promised incentive and no telephone option. Compared to Group 2 (4.4%), the response rate in Group 1 was almost twice as high (8.4%) at the end of the first data collection phase ($t(2{,}097) = 2.29$, $p < 0.05$). Of the respondents in Group 1 who were offered to complete the phone via the telephone, 26% responded by telephone in the early completion phase (phase 1). Turning to the incentives, 77 out of 361 sample members in Group 1 accepted the prepaid incentive (i.e., 21%), 66 of whom did so via PayPal and 11 via check. Fifty-seven of these 77 sample members accepted the prepaid incentive in phase 1 and eight of those 57 actually completed the survey in phase 1 (18 sample members completed in a later phase).

The production phase (phase 2) introduced the abbreviated interview in the aggressive data collection protocol for Group 1 only. Compared to Group 2 (12.1%)

who did not receive the offer to complete the abbreviated interview in this phase, Group 1 had a significantly higher response rate (22.7%; $t(2,097) = 3.67$, $p < 0.001$) although these results have to be interpreted with caution as Groups 1 and 2 are randomly assigned, but the results are conditional to Phase 1 outcomes..

Introducing the abbreviated survey at a later data collection stage (phase 3) for Group 2 did not seem to have the same effect on response rates we observed in phase 2 for Group 1 – but again, there is a dependency in phase 3 on what happened in the previous two phases.

The composition of the two NPSAS:16 FT respondent groups, that is, the base-year interview late respondents (Group 3) and the base-year interview early respondents (Group 4) was intentionally different in order to investigate whether less effort could be expended for individuals who completed the survey early. All statistical tests can therefore only serve a heuristic purpose. The base-year interview late respondents (Group 3) had a 70% response rate to the B&B:16/17 FT survey with the default protocol. The base-year interview early respondents (Group 4) responded to the follow-up survey at a 75% rate with a relaxed protocol. A naïve significance test between Group 3 and 4 shows that the response rates are significantly different at an alpha level of 0.05 ($t(2,097)=2.08$). This provides some evidence that a lower-cost effort for base-year early respondents as compared with base-year late respondents may still yield favorable results.

b. Field Test Design and Nonresponse Bias

Results.

Mahalanobis Distance & Nonresponse Bias. In addition to monitoring response rates, we calculated Mahalanobis distances to observe how the data collection protocols affected individual cases, although no resulting interventions were conducted during the field test due to the small sample size. Mahalanobis distances allowed us to identify sample members who were most likely to contribute to nonresponse bias. The input variables for the Mahalanobis distance modeling included institution characteristics (e.g., institutional sector) and characteristics for each sample member (e.g., sex and age).

Figure C.1 displays the average Mahalanobis distance over the course of the field test data collection period. After a sharp decline in the first 20 days of data collection, the average Mahalanobis distance only decreased marginally thereafter suggesting that that adding more sample members to the respondent pool did not necessarily contribute to a more representative sample.

Figure C.**1: Daily average Mahalanobis distance at the end of data collection**

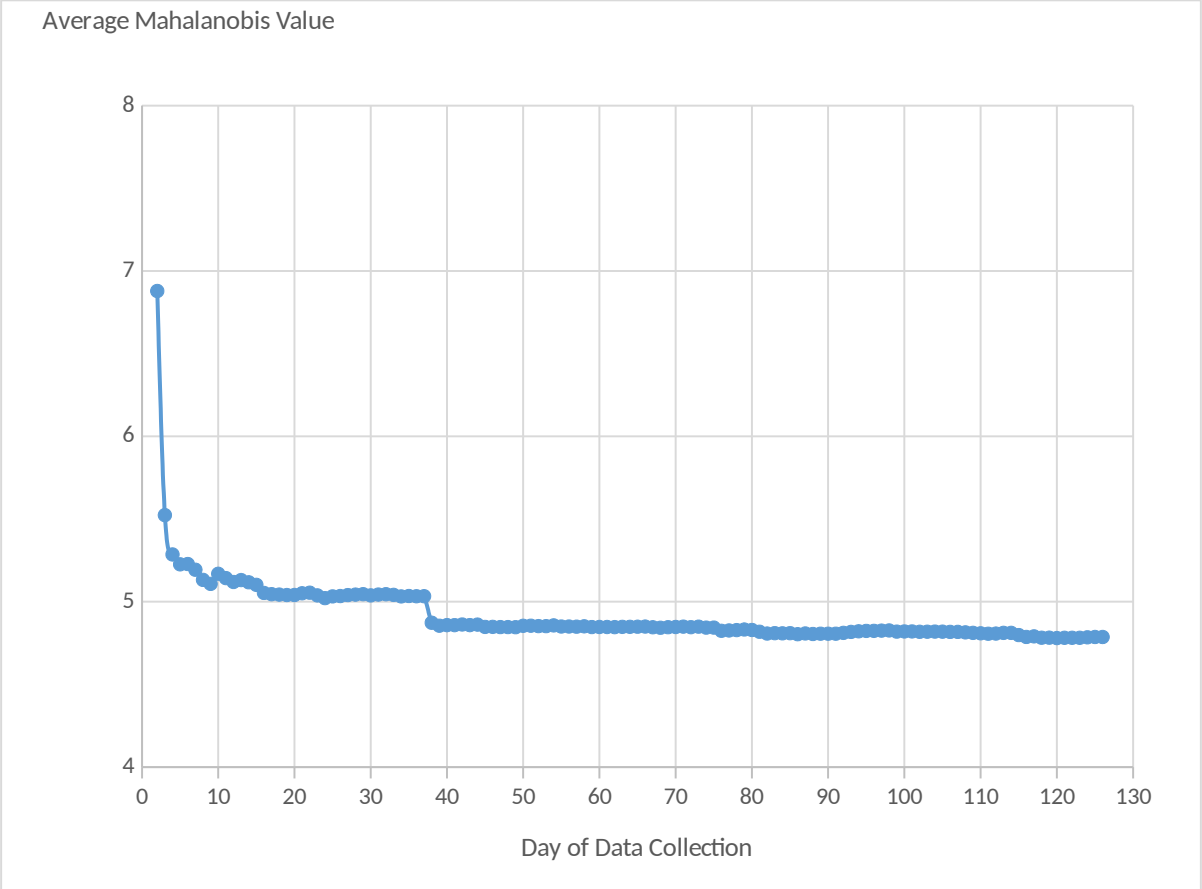
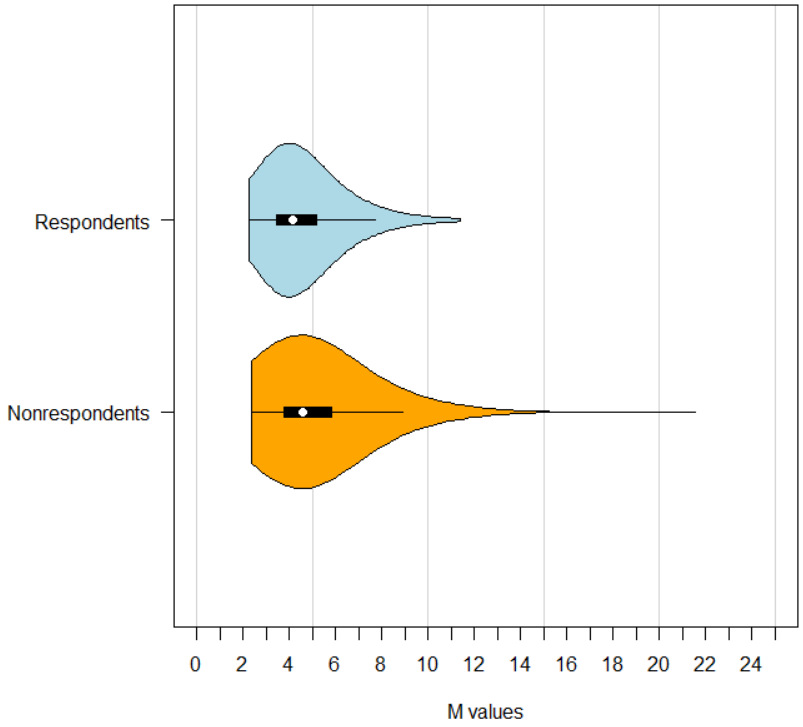


Figure C.2 displays the distribution of individual Mahalanobis distance measure for B&B:16/17 FT survey respondents (in blue) and B&B:16/17 FT survey nonrespondents (in orange). With the exception of one outlier, the distribution of both respondents and nonrespondents is very similar.

Figure C.**2:**

Distribution of Mahalanobis Distance Values by Response Status



Administrative Data and Nonresponse Bias. Based on administrative frame data, we conducted nonresponse bias analyses for sex, age, institutional sector of the NPSAS institution, region of the United States that the NPSAS institution is located in, and total enrollment counts. Table C.5 summarizes the results. In addition to significantly increasing response rates, the aggressive protocol seems to be contributing to less bias in the estimates (relative to the default nonrespondent protocol). While the average absolute relative bias in mean statistics was larger for the two nonresponse groups (Group 1 and Group 2) relative to the respondent groups (Group 3 and Group 4). We only found evidence of significant bias in Group 1 and this bias was present in only one of the 23 indicators we examined. These analyses confirm that overall, there is very little nonresponse bias in the examined estimates.

Table C.**5: Average absolute relative bias, median absolute relative bias, and percentage of significant deviations by data collection group**

Phase of B&B:16/17 FT	NPSAS:16 FT nonrespondents		NPSAS:16 FT respondents		
	Aggressive protocol (Group 1)	Default protocol (Group 2)	Late respondents-- default protocol (Group 3)	Early respondents-- relaxed protocol (Group 4)	Overall
Average absolute relative bias	14.30	18.33	4.87	3.98	5.64
Median absolute relative bias	9.98	14.29	2.66	2.57	4.51

c. Field Test Design and Differences in Measurement

Because we are bringing in potentially more reluctant respondents who might not be as conscientious in completing the survey and, for example, underreport the number of employers to avoid the extra burden associated with each reported employer, this might negatively impact data quality. As a final step, we hence compare whether the responses provided by respondents in the different data collection protocols differ due to potentially differential motivation to participate in the survey. Looking at these differences across data collection protocols[1], we did not see any statistically significant differences in reporting from respondents in the survey. The reported number of undergraduate and postbaccalaureate postsecondary institutions, employment, the number of employers, and whether respondents had any dependents did not differ when comparing Group 1 to Group 2 (see Table C.6), or when comparing Groups 1 and 2 to Groups 3 and 4 (see Table C.7). Because we are bringing in potentially more reluctant respondents, these results are reassuring in that increased data collection efforts in lower response propensity strata do not seem to decrease data quality.

Table C.**6: Test of total number of events, employment, and dependents among respondents in Group 1 and Group 2**

Item	NPSAS:16 FT nonrespondents		t-value / z-value*	p-value
	Aggressive protocol (Group 1)	Default protocol (Group 2)		
Number of undergraduate postsecondary institutions	0.53	0.76	-1.34	0.1856
Number of postbaccalaureate postsecondary institutions	0.33	0.29	0.41	0.6856
Any postbaccalaureate employment (ref. no)*	0.94	0.91	0.89	0.3744
Number of postbaccalaureate employers	1.45	1.56	-0.86	0.3923
Dependents (ref. no)*	0.03	0.12	-1.48	0.1380

[1] Comparing response distributions across groups, no evidence of differential nonresponse bias in each group was observed and hence no correction for differential selectivity was made.

Table C.7: Test of total number of events, employment, and dependents among respondents in Group 1 and Group 2 versus Group 3 and 4

Item	NPSAS:16 FT nonrespondents (Group 1 and 2)	NPSAS:16 FT respondents (Group 3 and 4)	t-value	p-value
Number of undergraduate postsecondary institutions	0.68	0.68	0.04	0.9682
Number of postbaccalaureate postsecondary institutions	0.31	0.26	0.89	0.3732
Any postbaccalaureate employment (ref. no)*	0.93	0.95	-1.15	0.2499
Number of postbaccalaureate employers	1.49	1.6	-1.51	0.1302
Dependents (ref. no)*	0.09	0.12	-1.01	0.3092

Recommendations for the full-scale study

Based on the increases in response rates with no negative effects on nonresponse bias, we recommend using the aggressive data collection protocol for the NPSAS:16 base-year nonrespondents and the default protocol for the NPSAS:16 late respondents with a minor modification in the B&B:16/17 full-scale data collection (see below). While the relaxed protocol worked relatively well for the NPSAS:16 field test early respondents (Group 4), we plan to modify this data collection protocol slightly in the full-scale study.

Based on what we learned in the B&B:16/17 FT, we propose the following modifications to the data collection protocols to be used in the full-scale study, to further increase response rates and reduce the potential for nonresponse error:

- We will not offer a prepaid incentive in the aggressive data collection group as only 21% of those offered the incentive accepted it, and only 34% of those completed the interview. Instead of offering a prepaid incentive, we recommend increasing the survey completion incentive for the NPSAS:16 non-located interview nonrespondents to $55, and increasing the survey completion incentive for the NPSAS:16 located interview nonrespondents and abbreviated respondents to $50. Base-year nonrespondents in the B&B:08/09 FT were offered $55 and had an overall response rate of 44%.
- Based on the low acceptance rate of the prepaid incentives, we instead propose to incentivize NPSAS:16 located interview nonrespondents and abbreviated respondents, and NPSAS:16 late respondents, with a $5 early bird incentive if they complete the survey within the first three weeks of data collection. Early bird or early response incentives have been shown to lead to faster responses and an increase in response rates and participation rates within the incentive period (e.g., LeClere et al., 2012; Coppersmith et al., 2016). This can increase efficiencies by reducing overall data collection cost and length of the field period for these cases.
- We learned that among base-year interview nonrespondents who participated in the B&B:16/17 FT survey, 22% were deemed ineligible by the survey. Base-year respondents had a 4.5% ineligibility rate. Therefore, we recommend sending an address update with an eligibility screener (tied to a $10 monetary

token of appreciation upon completion), to all sample members immediately upon OMB approval of the FS package. NPSAS non-study members will not receive the $10 for completion of the screener and will not be fielded in the survey even if found eligible in the screener. We propose to screen non-study members for eligibility so that we can remove them from the sample completely, if found to be ineligible.

- Compared to the B&B:08/09 FT response rates for base-year interview respondents (80.9%), the B&B:16/17 FT base-year interview response rates of Group 3 and Group 4 combined(72.6%) were considerably lower ($p < 0.001$). There are various possible reasons for the obvious decline in response rates, ranging from possible differences in locating efforts to a current trend in household surveys of declining response rates. Since we cannot disentangle the cause directly, we propose addressing declining response rates with increased incentives and light CATI interviewing. We recommend increasing incentives for the NPSAS:16 base-year early respondents (Group 4) to $30 to match those used in the default protocol (Group 3) and those used in NPSAS:16 FT, and to integrate light CATI interviewing with the relaxed protocol for Group 4.

C.2 Evaluation of Experiment #2: Questionnaire Design

The second experiment investigates potential sources of measurement error as a result of motivated underreporting by respondents who are asked to respond to a series of follow-up questions.

Respondents in surveys are often asked to respond to a series of follow-up questions that are repeated based on their response to filter questions (loops); for example, obtaining details about each employer a respondent has had. To determine the number of times a respondent goes through the loop, researchers can use one of two formats: (1) “how many” or (2) “go-again,” also sometimes referred to as the “grouped” or the “interleafed” formats respectively (Eckman et al., 2014). The “how many” format asks respondents to report the number of occurrences followed by questions asking details of each occurrence. The “go-again” format asks respondents to start with the first (or last) occurrence followed by more detailed questions. After answering the follow-up questions, respondents are asked if they have had any other occurrences. If “yes,” they continue to iterate through the loops. Such a task can become burdensome for respondents in either format, especially as the number of occurrences increases, potentially threatening data quality.

We expect to see that the reported number of occurrences will be lower in the “go-again” format as respondents learn that each additional occurrence triggers a set of follow-up questions. Thus, respondents in this format have a potentially reduced burden by underreporting the number of occurrences. Respondents in the “how many” format do not learn about the follow-up questions until after they report the number of events and hence have no prior knowledge of what is to follow. While the “how many” design should lead to higher reports of the number of events or occurrences, it might have adverse effects on data quality for the follow-up questions, because respondents might speed through the interview, provide a “don’t know” response, leave responses blank, or break off from the interview. Past

research supports evidence that the “how many” format provides more accurate responses for the filter question than the “go again” format, but the “go-again” format provides higher quality data for the follow-up questions (Eckman et al., 2014; Eckman and Kreuter, 2015).

We analyzed data from the B&B:16/17 FT where respondents were randomly assigned to one of the two loop formats asking about three areas: 1) number of undergraduate postsecondary institutions attended, 2) the number of postbaccalaureate postsecondary institutions attended, and 3) the number of postbaccalaureate employers and jobs. We evaluated the difference between loop formats in terms of number of reported occurrences, item nonresponse, response time, and breakoffs.

Respondents are randomly assigned to either the treatment group (“how many”) or the control group (“go-again”) at the start of the interview.

Table C.8 displays the number of cases assigned to the “go-again” (n=566) and “how many” (n=564) condition by final case disposition. Cases in each condition received different filter questions asking about undergraduate postsecondary institutions, postbaccalaureate postsecondary institutions, and postbaccalaureate employers and jobs but identical follow-up questions. Respondents who completed the abbreviated interview (abbreviated completed) received the section on postbaccalaureate employers but did not receive questions on postsecondary institutions. Respondents saw a maximum of seven loops per section (i.e., could report up to seven jobs within employers, for up to seven employers resulting in up to 49 jobs). Respondents who failed to provide the number of occurrences in the “how many” condition skipped the loop and were excluded from the following analyses (undergraduate postsecondary institutions n=5; postbaccalaureate postsecondary institutions n=1; postbaccalaureate employers n=10).

Table C.**8: Summary of experimental condition by disposition code**

	Loop experiment condition								
Case disposition	**Go-again**			**How many**			**Total**		
	N	Col %	Cum %	N	Col %	Cum %	N	Col %	Cum %
Final breakoff/partial	23	4.1	4.1	39	6.9	6.9	62	5.5	5.5
Abbreviated completed	69	12.2	16.3	77	13.7	20.6	146	12.9	18.4
Complete	474	83.7	100.0	448	79.4	100.0	922	81.6	100.0
Total	566	100.0		564	100.0		1130	100.0	

Note: Excluding pending partial cases.

Results. Consistent with earlier research (Eckman et al., 2014; Eckman and Kreuter, 2015) the results presented below suggest that data quality differs by loop format.

a. Number of reported occurrences

Table C.9 displays the distribution of responses to the filter questions for each topical section by experimental condition. Regardless of experimental condition, respondents were asked the follow-up questions if, and only if, they answered affirmatively to the filter question. For example, a respondent was only asked the

follow-up questions about undergraduate postsecondary institution if they first answered “yes” to the filter question asking if the respondent attended an undergraduate postsecondary institution. If respondents answered “yes”, those assigned to the “how many” condition were then asked to report the total number of undergraduate postsecondary institutions attended before beginning the follow-up questions, whereas the “go-again” group transitioned immediately from the filter to the follow-up questions. The distribution of responses did not differ by experimental condition for any of the topical sections (undergraduate postsecondary institutions z = -0.16, p = 0.87; postbaccalaureate postsecondary institutions z = 0.64, p = 0.52; postbaccalaureate employers z = -0.39, p = 0.70).

Table C.9: Responses by section and experimental condition

	Loop experiment condition								
Loop Section	**Go-again**			**How many**			**Total**		
	N	Col %	Cum %	N	Col %	Cum %	N	Col %	Cum %
Undergraduate postsecondary institutions yes/no									
No	271	47.9	47.9	257	45.6	45.6	528	46.7	46.7
Yes	221	39.0	86.9	214	37.9	83.5	435	38.5	85.2
Missing (yes/no and how many)	1	0.2	87.1	6	1.1	84.6	7	0.6	85.8
Abbreviated	73	12.9	100.0	87	15.4	100.0	160	14.2	100.0
Total	566	100.0		564	100.0		1130	100.0	
Postbaccalaureate postsecondary institutions yes/no									
No	363	64.1	64.1	359	63.7	63.7	722	63.9	63.9
Yes	130	23.0	87.1	117	20.7	84.4	247	21.9	85.8
Missing (yes/no and how many)	0	0.0	87.1	1	0.2	84.6	1	0.1	85.8
Abbreviated	73	12.9	100.0	87	15.4	100.0	160	14.2	100.0
Total	566	100.0		564	100.0		1,130	100.0	
Postbaccalaureate employers and jobs yes/no									
No	34	6.0	6.0	30	5.3	5.3	64	5.7	5.7
Yes	532	94.0	100.0	519	92.0	97.3	1051	93.0	98.7
Missing (yes/no and how many)	0	0.0	100.0	15	2.7	100.0	15	1.3	100.0
Total	566	100.0		564	100.0		1,130	100.0	

Note: Respondents who did not provide information on the number of occurrences in the how many condition (i.e., the number of institutions, employers or jobs) skipped the loop experiment and are captured in the missing category.

Table C.10 presents the total number of occurrences reported in each topical section by experimental condition.[2] The total number of occurrences in each section corresponds to the reported number of occurrences in the “how-many” condition and to the sum of affirmative responses to the filter questions in the “go-again” condition after each loop. There was no difference in the average number of undergraduate postsecondary institutions reported by respondents in the “how many” (x-bar = 0.72) and “go-again” (x-bar = 0.64) conditions (t(926.37)=-1.44, p

[2] Since we are dealing with count data, we replicated all tests using a Negative Binomial regression (the assumptions for a more parsimonious Poisson regression do not hold). The results are identical.

= 0.15). There was also no difference in the average number of postbaccalaureate postsecondary institutions section between the “how many” (x-bar = 0.26) and the “go-again” condition (x-bar = 0.27) in the ($t(967) = 0.51$, $p = .61$) is also nonsignificant.

The average number of reported postbaccalaureate employers and jobs is higher in the “how many” condition compared to the “go-again” condition. More specifically, respondents in the “go-again” condition report on average 1.38 employers with an average of 1.51 jobs across all employers and 1.1 jobs per employer. In the “how many” condition, respondents report 1.79 employers with 2.27 jobs across all employers and 1.32 jobs per employer. The differences in these reports are statistically significant (employers: $t(972.25) = -7.35$, $p < 0.001$; jobs across all employers: $t(757.53) = -8.82$, $p < 0.001$; jobs per employer: $t(602.63)=-5.43$; , $p < 0.001$) and are driven by respondents who report two or more employers.[3]

Table C.**10: Test of total number of events reported in each section by experimental condition**

Loop Section	Go-again	How many	t-value	p-value	N
Undergraduate postsecondary institutions	0.64	0.72	-1.44	0.1493	963
Postbaccalaureate postsecondary institutions	0.27	0.26	0.51	0.6075	969
Postbaccalaureate employers (censored)	1.38	1.79	-7.35***	0.0001	1,115
Postbaccalaureate jobs (censored)+	1.51	2.27	-8.82***	0.0001	1,088

* p<0.05, ** p<0.01, ***p<0.001.
Note: Based on respondents who completed the loop sections (for breakoff analyses see below). The number of employers was censored at 7 (n=1) and the number of jobs was censored at 12 (n=2).
+The number of postbaccalaureate jobs is the average number of jobs across employers.

b. Completion time

We used time stamps capturing the total time and the load time to derive the average response time spent on each screen within each section. Response time measures the time that the respondent sees the screen after it was loaded (in seconds). The load time refers to the time the website takes to load (in seconds). The total time is the sum of the two (in seconds). Due to backups and timeouts some time information was missing. These cases were deleted from the analyses before aggregation (by variable and not listwise) and outliers were censored (by time measure). Thus, response times and load times may not add up to total time. In order to compare completion times across experimental conditions, all time data were divided by the number of screens a respondent saw, accounting for the total number of loops. Table C.11 displays the number of cases for whom timing information is available.

Table C.**11: Cases with available timing information by experimental condition**

[3] Subsetting the analyses by respondents who reported only one or fewer employers shows that the difference is no longer significant ($t(502.32) = 1.02$, $p = 0.31$) whereas the difference among respondents who report two or more employers is statistically significant ($t(487.02) = -5.02$, $p < 0.001$).

	Loop experiment condition					
Loop Section	**Go-again**		**How many**		**Total**	
	N	Row %	N	Row %	N	Row %
Undergraduate postsecondary institutions	221	50.8	214	49.2	435	100.0
Postbaccalaureate postsecondary institutions	130	52.6	117	47.4	247	100.0
Postbaccalaureate employers and jobs	531	52.3	484	47.7	1,015	100.0

Note: Based on respondents who completed the loop sections. Backups and timeouts caused negative and missing time stamps (especially in the "how many" condition in the postbaccalaureate employers and jobs section). These cases were deleted from the analyses (undergraduate postsecondary institutions n=0; postbaccalaureate postsecondary institutions n=0; postbaccalaureate employers "go-again" n=1, "how many" = 35). Outliers were identified and censored at the page level for each topical section using a log-transformation +/- 1.5 * interquartile range from the median (undergraduate postsecondary institutions 5.81%; postbaccalaureate postsecondary institutions 6.04%; postbaccalaureate employers 9.62%).

Table C.12 provides the results of testing for differences in the average response time, load time, and total time per screen across the experimental conditions (average times are displayed in seconds; the t-test is based on the log-transformed data per page). Overall, response, load, and total time did not differ within the undergraduate or the postbaccalaureate postsecondary institutions section.

In the postbaccalaureate employers and jobs section, respondents in the "how many" condition spent approximately half a second less on each screen (11.40 sec.) compared to the "go-again" condition (11.96 sec.). This difference is statistically significant ($t(875.76) = 3.47$, $p < 0.001$).. Load time was significantly longer in the "how many" condition compared to the "go-again" condition ($t(974.72) = -3.36$, $p < 0.001$) although this difference is marginal (0.82 sec. vs 0.99 sec.). Longer loading times in the "how many" condition are plausible due to a more complex skip logic and internal routing. As a result of these counteracting trends, the difference in total time is decreased, albeit still significant ($t(940.49) = 2.56$, $p < 0.05$). Respondents in the "go-again" condition spent on average 12.87 seconds on each form whereas respondents in the "how many" condition spent on average 12.44 seconds per screen.

Table C.**12: Test of mean time to complete a screen in each section by experimental condition (in seconds)**

Loop Section	**Go-again**	**How many**	**t-value**	**p-value**
Undergraduate postsecondary institutions (n=435)				
Response time per screen	12.55	12.87	-0.61	0.5453
Load time per screen	0.57	0.60	-0.34	0.7350
Total time per screen	13.17	13.47	-0.63	0.5289
Postbaccalaureate postsecondary institutions (n=247)				
Response time per screen	13.01	13.65	-0.73	0.4686
Load time per screen	0.71	0.70	0.05	0.9563
Total time per screen	13.73	14.34	-0.91	0.3625
Postbaccalaureate employers and jobs (n=1,015)				
Response time per screen	11.96	11.40	3.42***	0.0007
Load time per screen	0.82	0.99	-3.34***	0.0009
Total time per screen	12.87	12.44	2.54*	0.0113

* p<0.05, ** p<0.01, ***p<0.001.

Note: Based on respondents who completed the loop sections. Backups and timeouts caused

negative time and missing time stamps. Outliers were censored. This is why response and load times may not add up to total time. T-tests are based on the log-transformed data per form.

Table C.13 provides the results of testing the differences in the average total time respondents took to complete each loop section by experimental condition (in minutes). Average total completion time did not differ within the undergraduate or the postbaccalaureate postsecondary institutions section.

Respondents in the “how many” condition (12.87 min.) took approximately 2 minutes longer to complete the postbaccalaureate employers and jobs section compared to the “go-again” condition (10.87 min.). This difference is statistically significant ($t(978.80)=-3.15$, $p < 0.01$). This increased length is due to the significantly higher number of reported employers and jobs in the “how many” condition compared to the “go-again” condition (see Section C2. a. Number of reported occurrences).

Table C.13: Test of mean total time to complete each section by experimental condition (in minutes)

Loop Section	Go-again	How many	t-value	p-value
Undergraduate postsecondary institutions (n=435)	3.12	3.38	0.02	0.9843
Postbaccalaureate postsecondary institutions (n=247)	2.99	2.86	0.76	0.4499
Postbaccalaureate employers and jobs (n=1,015)	10.87	12.59	-3.15*	0.0017

* p<0.05, ** p<0.01, ***p<0.001.

Note: Based on respondents who completed the loop sections. Backups and timeouts caused negative time and missing time stamps. Outliers were censored. This is why response and load times may not add up to total time. T-tests are based on the log-transformed data per form.

c. Item nonresponse

Measuring item nonresponse is complicated by the fact that respondents in the “go-again” condition who fail to report another institution or employer will by design not have any item nonresponse for these missing occurrences. Respondents in the “how many” condition are presented with the number of loops corresponding to the number of occurrences reported and hence have an increased number of opportunities to skip items. Thus, by design, item nonresponse is expected to be significantly higher in the “how many” condition. As a result, we investigated whether a respondent in either condition ever failed to respond to a follow-up question.

The same pattern as for the substantive responses and the completion time emerges. Item nonresponse does not differ by experimental condition in the undergraduate and postbaccalaureate postsecondary institutions topical sections (see Table C.14). In contrast, more respondents in the “how many” condition (64.7%) skipped at least one item compared to respondents in the “go-again” condition (53.0%) in the postbaccalaureate employers and jobs section ($z= -3.86$, $p < 0.001$).

Table C.14: Test of item nonresponse in each section by

experimental condition (in percent)					
Loop Section	**Go-again**	**How many**	**z-value**	**p-value**	**N**
Undergraduate postsecondary institutions	0.050	0.061	-0.50	0.6163	435
Postbaccalaureate postsecondary institutions	0.108	0.103	0.13	0.8957	247
Postbaccalaureate employers and jobs	0.530	0.647	-3.86***	0.0001	1,051

* p<0.05, ** p<0.01, ***p<0.001.
Note: Based on respondents who started the loop sections.

d. Breakoff rates

The breakoff analyses investigated whether or not a respondent broke off during the survey. Again, we do not see any significant differences in the breakoff rates by experimental condition in the undergraduate and postbaccalaureate postsecondary institution topical sections (see Table C.15). Breakoff rates are significantly higher (z= -1.98, p < 0.05) in the "how many" condition (15.2 percent) compared to the "go-again" condition (11.1 percent) in the postbaccalaureate employers and jobs section.

Table C.**15: Test of breakoff rates in each section by experimental condition (in percent)**

Loop Section	**Go-again**	**How many**	**z-value**	**p-value**	**N**
Undergraduate postsecondary institutions	0.032	0.047	-0.81	0.4179	435
Postbaccalaureate postsecondary institutions	0.008	0.000	0.95	0.3418	247
Postbaccalaureate employers and jobs	0.111	0.152	-1.98*	0.0474	1,051

* p<0.05, ** p<0.01, *** p<0.001.
Note: Based on respondents who started the loop sections.

Recommendations for the full-scale study

The two question formats did not yield significantly different estimates in the undergraduate and the postbaccalaureate postsecondary institutions section although there is a significant difference in the number of undergraduate postsecondary institutions when investigating only respondents who reported at least two undergraduate postsecondary institutions ("go again" = 2.31, "how many" = 2.49; t(153.73) = -1.70, p <0.10). The "how many" condition yielded significantly higher estimates in the postbaccalaureate employment and jobs section. Further analysis showed that these results are driven by those respondents reporting two or more employers. These results are plausible, as only those respondents who experience multiple loops can potentially "learn" that reporting another institution or employer yields more follow-up questions and that reporting fewer conditions in the "go-again" condition might avoid additional burden. The results by Eckman and Kreuter (2015) support this explanation.

While there is no significant difference in item nonresponse or breakoff rates in the undergraduate and the postbaccalaureate postsecondary institutions section, item nonresponse and breakoff rates are significantly higher in the postbaccalaureate employment and jobs section. One possible explanation is that the “how many” condition forces respondents through more loops ;hence, increases respondent burden which is known to be associated with item nonresponse and breakoff. Another possible explanation is that by design, the item nonresponse and breakoff rates in the “go-again” condition are underestimated as we do not know what they would have been, had the respondents not failed to report an occurrence. As we do not have information about the counterfactual regarding item nonresponse and breakoffs in the “go again” condition and don’t know in which loop the breakoff occurred in the “how many” condition, we cannot differentiate between these scenarios.

Despite the marginal loss in data quality in the follow-up questions, we recommend using the “how many”
format in the B&B:16/17 full-scale data collection. The increased accuracy in the estimated number of occurrences in the “how many” condition provides critical information even without the follow-up questions, and allows researchers to better impute missing information as it is obvious which information is missing.

References

Coppersmith, J., Vogel, L.K., Bruursema, T. and K. Feeney. 2016. Effects of Incentive Amount and Type of Web Survey Response Rates. Survey Practice, no pp.

Eckman, S. and F. Kreuter. 2015. Misreporting to Looping Questions in Surveys. Recall, Motivation, and Burden. IAB Discussion Paper 29/2015.

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Presser, S., and R. Tourangeau. 2014. Assessing Mechanisms of Misreporting to Filter Questions. Public Opinion Quarterly, 78(3): 721-733.

LeClere, F., Plumme, S., Vanicek, J., Amaya, A. and K. Carris. 2012. Household Early Bird Incentives: Leveraging Family Influence to Improve Household Response Rates.” American Statistical Association Joint Statistical Meetings, Section on Survey Research.