

2020/25 Beginning Postsecondary Students (BPS:20/25) Full-Scale Study

Appendix D

Results of the BPS:20/25 Field Test Experiments

OMB # 1850-0631 v.21

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

October 2024

Contents

Results of the BPS:20/25 Field Test Experiments.....	2
D.1 Address Confirmation Experiment.....	3
Results.....	3
Recommendations for the full-scale study.....	5
D.2 Telephone Reminder Experiment.....	5
Results.....	5
Recommendations for the full-scale study.....	8

Tables

Table 1. Overview of indicators, operationalizations and analytic approaches for data collection experiments.....	2
Table 2. Response rates by address confirmation incentive experimental condition.....	3
Table 3. Sample composition by address confirmation incentive experimental condition	4
Table.4. Average number of days to complete by address confirmation incentive experimental condition.....	5
Table 5. Response rates by telephone reminder experimental condition.....	5
Table 6. Sample composition by reminder mode experimental condition.....	7
Table 7. Average number of days to complete by reminder mode experimental condition and evaluation period.....	8

Results of the BPS:20/25 Field Test Experiments

The BPS:20/25 field test included two data collection experiments focused on survey participation. The results of these field test experiments are summarized below. Full details of the experiments were described and approved in BPS:20/25 Field Test Supporting Statement Part B (OMB# 1850-0631 v.20).

The data collection experiments explored the effectiveness of:

- 1) an “address confirmation” incentive where respondents received an additional \$5 if they provided their phone number, mailing or e-mail address before the start of the survey (i.e., the “address confirmation questions”), and
- 2) eliminating telephone reminders for sample members who responded to NPSAS:20 and BPS:20/22.

The experiments were evaluated on three criteria: survey response, sample representativeness, and data collection efficiency. Survey response was evaluated for both experiments using response rates. Pearson chi-squared tests assessed whether the experimental treatment significantly increased survey response.

Next, using administrative frame data, sample representativeness was assessed across sex, race, ethnicity, and control of institution (i.e., public institution, private nonprofit institution, and private for-profit institution). Pearson chi-squared tests were used to compare the distribution of respondent characteristics (i.e., percentages for sex, race, ethnicity, and institutional control) across the treatment and control groups. Nonresponse bias was also calculated (Groves et al. 2009) for each representativeness metric within each group (i.e., within control and experimental groups) as

$$\frac{m}{n}(\bar{y}_r - \bar{y}_m)$$

where m is the number of nonrespondents in a group, n is the total sample size for that group, \bar{y}_r is the respondent mean for that group, and \bar{y}_m is the nonrespondent mean for that group. This provides a measure of the magnitude of bias. To compare the magnitude of biases across the control and treatment group, the difference of absolute biases was calculated as

$$|Bias_T| - |Bias_C|$$

where $|Bias_T|$ is the absolute value of nonresponse bias for the treatment group and $|Bias_C|$ is the absolute value of the control group.

Finally, data collection efficiency is operationalized as the number of days between the start of the experiment and survey completion. One-sided t -tests were used to explore whether the experimental treatments significantly reduced the number of days from the start of the experiment it took respondents to complete the survey. Table 1 summarizes these indicators, their operationalization, and the analytic approaches.

Table 1. Overview of indicators, operationalizations and analytic approaches for data collection experiments

Indicator	Operationalization	Analytic Approach
Survey response	Response rates (eligible sample members only, including partial completions)	X^2 test

Sample representativeness	Nonresponse bias	$\frac{m}{n}(\bar{y}_r - \bar{y}_m)$
Data collection efficiency	Rate of respondent characteristics Number of days from start of experiment to survey completion	χ^2 test t -test

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/25 Beginning Postsecondary Students (BPS:20/25) Field Test.

D.1 Address Confirmation Experiment

Feedback from sample members in open-ended questions on NPSAS:20 and BPS:20/22 as well as focus group participants have noted that the authenticity of the BPS survey request is sometimes questioned. Sample members question the legitimacy of a survey (and by extension, the legitimacy of a post-paid incentive offer for completing a survey) are more likely to ignore requests for survey participation (Dillman et al. 2014), thus decreasing response rates and increasing the potential for nonresponse bias (Massey and Tourangeau 2012). The “address confirmation” experiment investigated the effectiveness of giving respondents an additional \$5 incentive (treatment group; $n = 1,330$) if, at the beginning of the survey, they provided their contact information (i.e., phone number or mailing or e-mail address) via several address confirmation questions. This incentive was designed to establish the authenticity of the survey request: providing sample members with a small up-front incentive for answering address questions may demonstrate that the survey is legitimate, theoretically motivating hesitant sample members to complete. The control group was also asked to provide the same contact information but was not offered an additional incentive ($n = 1,331$).

Results.

Survey response. At the end of data collection (17 weeks in the field), the overall survey response rate for the \$5 address confirmation treatment group (61.9 percent) was significantly lower than the control group (67.2 percent; $\chi^2 = 8.36, p < .01$; first row of Table 2).

Response rates across experimental conditions were also examined within the two data collection groups: 1) the default data collection group (i.e., sample members who responded to NPSAS:20 and BPS:20), and 2) the aggressive data collection group (i.e., sample members who were NPSAS:20 survey nonrespondents, BPS:20/22 survey nonrespondents, or BPS:20/22 survey final partials). In the default data collection group, survey response rates for the \$5 address confirmation treatment group (70.6 percent) was marginally significantly lower than the control group 74.5 percent; $\chi^2 = 3.82, p = .05$; second row of Table 2. In the aggressive data collection group, survey response rates for the \$5 address confirmation treatment group (35.9 percent) was significantly lower than the control group (46.0 percent); $\chi^2 = 7.08, p < .01$; third row of Table 2.

The results from this study indicate that a \$5 incentive was not effective at encouraging survey response, and surprisingly, it seems that this incentive offer discouraged response to the survey. This effect was true overall and within both data collection groups, though it was more pronounced in the aggressive data collection group. Although sample members in the experiment’s treatment group were offered a higher total incentive than the control group (baseline plus \$5), it is possible that an

offer of an extra \$5 incentive was confusing to sample members. They may have interpreted the \$5 offer to be the *only* incentive available for the survey rather than as an *additional* incentive, thereby discouraging response versus an offer of the baseline incentive alone. Alternatively, highlighting two steps required to complete the survey in the treatment condition (where the contacting materials asked sample members to complete the address completion questions, and then complete the survey) may have induced a perception of higher burden for the treatment group versus the contact group (where the contacting materials only asked participants to complete the survey itself; the address confirmation questions were not mentioned).

Table 2. Response rates by address confirmation incentive experimental condition, overall and by data collection group

	Control Group No incentive for address confirmation	Treatment Group \$5 for address confirmation	χ^2	<i>p</i> -value
Overall response rate	67.2	61.9	8.36	0.004
Default data collection group response rate	74.5	70.6	3.82	0.051
Aggressive data collection group response rate	46.0	35.9	7.08	0.008

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/25 Beginning Postsecondary Students (BPS:20/25) Field Test.

Representativeness. Sample representativeness of respondents was investigated at the end of data collection. The percentage of BPS:20/25 field test respondents who identified as female did not significantly differ across the no incentive control group (61.2 percent) and the \$5 incentive treatment group (59.4 percent; $\chi^2 = 0.61$, $p = 0.43$). Similarly, the percentage of White respondents in the control group (59.3 percent) was not significantly different from the percentage of White respondents in the treatment group (61.1 percent; $\chi^2 = 0.60$, $p = 0.44$), nor was there a statistically significant difference between the percentage of Hispanic or Latino respondents in the control group (20.1 percent) and the treatment group (21.2 percent; $\chi^2 = 0.89$, $p = 0.35$) below displays sample composition for each of respondent characteristic by experimental condition.

The percentage of BPS:20/25 field test respondents who identified as female did not significantly differ across the no incentive control group (61.2 percent) and the \$5 incentive treatment group (59.4 percent; $\chi^2 = 0.61$, $p = 0.43$). Similarly, the percentage of White respondents in the control group (59.3 percent) was not significantly different from the percentage of White respondents in the treatment group (61.1 percent; $\chi^2 = 0.60$, $p = 0.44$), nor was there a statistically significant difference between the percentage of Hispanic or Latino respondents in the control group (20.1 percent) and the treatment group (21.2 percent; $\chi^2 = 0.89$, $p = 0.35$)

There was also no significant difference between the percentage of respondents from public institutions in either the control group (58.3 percent) or the treatment group (60.5 percent; $\chi^2 = 0.85$, $p = 0.36$). The same was also true for respondents from private nonprofit institutions ($\chi^2 = 0.29$, $p = 0.59$) and private for-profit institutions ($\chi^2 = 0.39$, $p = 0.53$).

Nonresponse bias tended to be lower in the \$5 incentive treatment group (last column of table 3), with four of the six indicators demonstrating smaller biases with the treatment group (ranging from 0.11 to 1.37). However, the magnitudes of these biases were similar across the control and treatment groups for all indicators (average difference of absolute bias across the six indicators = 0.98).

Taken together, these results indicate that offering a \$5 incentive for completing the address confirmation questions encouraged response from the same types of sample members as did the baseline incentive alone.

Table 3. Sample composition by address confirmation incentive experimental condition

Sample member characteristic	Treatment Group		Test of independence for sample characteristic across treatment and control group respondents		Difference in Bias Treatment - Control
	Control Group No incentive for address confirmation	Treatment Group \$5 for address confirmation	χ^2	p-value	
Female (in percent)					
Respondents	61.2	59.4	0.61	0.43	
Nonrespondents	52.8	56.4			
Respondents – nonrespondents			3.0		
Nonresponse bias	2.76	1.14			
Overall sample (n = 2,630) ¹	58.5	58.3			
White (in percent)					
Respondents	59.3	61.1	0.60	0.44	
Nonrespondents	64.4	60.5			
Respondents – nonrespondents	-5.1	0.6			
Nonresponse bias	-1.67	0.23			
Overall sample (n = 2,580) ¹	61.0	60.9			
Hispanic or Latino (in percent)					
Respondents	20.1	21.1	0.89	0.35	
Nonrespondents	22.8	22.9			
Respondents – nonrespondents	-2.7	-1.7			
Nonresponse bias	-0.88	-0.64			
Overall sample (n = 2,600) ¹	21.0	22.3			
Control of institution (in percent)					
Public					
Respondents	58.3	60.5	0.85	0.36	
Nonrespondents	61.2	57.2			
Respondents – nonrespondents	-2.9	3.3			
Nonresponse bias	-0.95	1.26			
Overall sample (n = 2,660) ¹	59.3	59.3			
Private nonprofit					
Respondents	28.3	27.1	0.29	0.59	
Nonrespondents	24.5	26.8			
Respondents – nonrespondents		0.3			
Nonresponse bias	1.24	0.11			
Overall sample (n = 2,660) ¹	27.1	27.0			
Private for-profit					
Respondents	13.4	12.4	0.39	0.53	
Nonrespondents	14.2	16.0			
Respondents – nonrespondents	-0.8	-3.6			
Nonresponse bias	-0.26	-1.37			
Overall sample (n = 2,660) ¹	13.7	13.8			

¹ Sample sizes for the overall sample differ due to missing data.

NOTE: Nonresponse bias is calculated as (respondent mean - nonrespondent mean) * (nonresponse rate).

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/25 Beginning Postsecondary Students Longitudinal Study (BPS:20/25) Field Test.

Efficiency. Data collection efficiency was examined (i.e., the average number of days it took respondents to complete the survey) at the end of data collection (17 weeks in the field). Table.4 summarizes the average number of days to survey completion for the control and \$5 incentive treatment groups.

At the end of data collection, respondents in the treatment group took about the same number of days (36.7 days) as respondents in the control group (35.9 days) to complete the survey ($t(1, 1,677) = 0.47, p = 0.64$). This indicates that the \$5 treatment for completing the address confirmation questions did not encourage respondents to complete their surveys any faster than respondents in the control group.

Table.4. Average number of days to complete by address confirmation incentive experimental condition

	Control Group No incentive for address confirmation	Treatment Group \$5 for address confirmation	<i>t</i>	<i>p</i> -value
End of data collection	35.9	36.7	0.47	0.64

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/25 Beginning Postsecondary Students (BPS:20/25) Field Test.

Recommendations for the full-scale study

Offering sample members \$5 to complete the address confirmation questions in addition to the baseline post-paid incentive amount did not significantly improve representativeness or days to complete by the end of data collection. Further, the \$5 treatment actually depressed overall survey response rates. Therefore, the use of a \$5 address confirmation incentive in the BPS:20/25 full-scale data collection is not recommended.

D.2 Telephone Reminder Experiment

Although telephone reminders are commonly used as a data collection strategy for BPS sample members, this method is more expensive than other reminder methods (e.g., text and e-mail reminders). Further, the percentage of respondents who complete their survey using telephone is low (4 percent in NPSAS:20; 4 percent in BPS:20/22 field test; 5 percent in BPS:20/22 full-scale). The second data collection experiment examined the implications of eliminating telephone reminders to sample members who were deemed likely to respond to the BPS:20/25 field test based on their past response behavior (i.e., sample members who responded to NPSAS:20 and BPS:20/22, known as the default data collection group). Eight weeks into data collection, nonresponding sample members in the default data collection group ($n = 930$) were randomly assigned to either 1) receive telephone prompting as in past BPS administrations (the control group; $n = 466$), or 2) not receive any telephone prompting (the treatment group; $n = 464$). All other data collection activities (e.g., reminder e-mails, hardcopy mailings) continued for both groups.

Results.

Survey response. Response rates were examined at the end of data collection (17 weeks in the field). Table 5 summarizes the response rates for the telephone reminder control group and the no telephone reminder experimental group.

At the end of the reminder period, the response rate for the no telephone reminder treatment group (33.8 percent) was 14.9 percentage points lower the control group who did receive telephone reminders (48.7 percent), a significant difference ($\chi^2 = 21.22, p < 0.001$). This indicates that removing telephone reminder significantly reduced the response rate of sample members included in the experiment (i.e., among those who had not responded by week 8 of data collection).

Table 5. Response rates by telephone reminder experimental condition

	Control Group Telephone Reminders	Treatment Group No Telephone Reminders	χ^2	p-value
End of data collection	48.7	33.8	21.22	< 0.001

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/25 Beginning Postsecondary Students (BPS:20/25) Field Test.

Representativeness. Sample representativeness of respondents was investigated at the end of data collection. The percentage of BPS:20/25 field test respondents who identified as female did not significantly differ across the no incentive control group (61.2 percent) and the \$5 incentive treatment group (59.4 percent; $\chi^2 = 0.61$, $p = 0.43$). Similarly, the percentage of White respondents in the control group (59.3 percent) was not significantly different from the percentage of White respondents in the treatment group (61.1 percent; $\chi^2 = 0.60$, $p = 0.44$), nor was there a statistically significant difference between the percentage of Hispanic or Latino respondents in the control group (20.1 percent) and the treatment group (21.2 percent; $\chi^2 = 0.89$, $p = 0.35$) below displays sample composition for each of respondent characteristic by experimental condition.

The percentage of BPS:20/25 field test respondents who identified as female did not significantly differ across the telephone reminder control group respondents (56.9 percent) and the no telephone reminder treatment group (55.4 percent; $\chi^2 = 0.08$, $p = 0.78$). Similarly, the percentage of White respondents in the control group (63.1 percent) was not significantly different from the percentage of White respondents in the treatment group (65.6 percent; $\chi^2 = 0.26$, $p = 0.61$), nor was there a statistically significant difference between the percentage of Hispanic or Latino respondents in the control group (21.9 percent) and the treatment group (18.6 percent; $\chi^2 = 0.61$, $p = 0.44$)

There was also no significant difference between the percentage of respondents from public institutions in either the control group (56.8 percent) or the treatment group (59.9 percent; $\chi^2 = 0.35$, $p = 0.55$). The same was also true for respondents from private nonprofit institutions ($\chi^2 = 0.35$, $p = 0.56$) and private for-profit institutions ($\chi^2 = 0.01$, $p = 0.93$).

Nonresponse bias tended to be lower in the telephone reminder control group (last column of table 6), with five of the six indicators demonstrating smaller absolute biases with the control group (ranging from 0.1 to 0.9). However, the magnitudes of these biases were similar across the control and treatment groups for all indicators (average difference of absolute bias across the six indicators = 1.28).

Taken together, these results indicate that not offering telephone reminders encouraged response from the same types of sample members as did offering a telephone reminder.

Table 6. Sample composition by reminder mode experimental condition

Sample member characteristic	Distribution of characteristic across treatment and control group respondents		χ^2	p-value	Difference in Bias Treatment - Control
	Control Group Telephone reminder	Treatment Group No telephone reminder			
Female (in percent)					
Respondents		56.9	55.4	0.08	
Nonrespondents		56.4	54.1		
Respondents – nonrespondents			1.3		
Nonresponse bias			0.9		0.6
Overall sample (n = 920) ¹		56.7	54.6		
White (in percent)					
Respondents		63.1	65.0	0.26	
Nonrespondents		63.2	66.1		
Respondents – nonrespondents		-0.1	-0.5		
Nonresponse bias		-0.1	-0.3		0.2
Overall sample (n = 910) ¹		63.1	65.9		
Hispanic or Latino (in percent)					
Respondents		21.9	18.0	0.61	
Nonrespondents		21.6	22.6		
Respondents – nonrespondents			-4.0		
Nonresponse bias			-2.6		2.4
Overall sample (n = 910) ¹		21.7	21.2		
Control of institution (in percent)					
Public					
Respondents		56.8	59.0	0.35	
Nonrespondents		58.6	56.0		
Respondents – nonrespondents		-1.8	3.9		
Nonresponse bias		-0.9	2.6		1.7
Overall sample (n = 930) ¹		57.7	57.0		
Private nonprofit					
Respondents		28.2	25.0	0.35	
Nonrespondents		27.6	30.0		
Respondents – nonrespondents			-4.5		
Nonresponse bias			-3.0		2.7
Overall sample (n = 930) ¹		27.9	28.5		
Private for-profit					
Respondents		15.0	14.0	0.01	
Nonrespondents		13.8	14.0		
Respondents – nonrespondents			0.7		
Nonresponse bias			0.5		-0.1
Overall sample (n = 930) ¹		14.4	14.2		

¹ Sample sizes for the overall sample differ due to missing data.

NOTE: Nonresponse bias is calculated as (respondent mean - nonrespondent mean) * (nonresponse rate).

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/25 Beginning Postsecondary Students Longitudinal Study (BPS:20/25) Field Test.

Efficiency. Data collection efficiency was examined (i.e., the average number of days it took respondents to complete the survey) at the end of data collection (17 weeks in the field). At the end of data collection, respondents in the treatment group took longer (89.7 days) than respondents in the control group (83.6 days) to complete the survey ($t(1, 363) = -3.83, p < 0.001$). However, this difference is small (less than a week; 6.1 days), and not long enough to allow for any cost savings in the data collection process (e.g., via fewer reminder calls, texts, or mailings). Thus, while these differences are statistically significant, they are not practically significant. summarizes the average number of days to survey completion for the control and treatment groups at these two time points.

At the end of data collection, respondents in the treatment group took longer (89.7 days) than respondents in the control group (83.6 days) to complete the survey ($t(1, 363) = -3.83, p < 0.001$). However, this difference is small (less than a week; 6.1 days), and not long enough to allow for any cost savings in the data collection process (e.g., via fewer reminder calls, texts, or mailings). Thus, while these differences are statistically significant, they are not practically significant.

Table 7. Average number of days to complete by reminder mode experimental condition and evaluation period

	Control Group Telephone reminder	Treatment Group No telephone reminder	<i>t</i>	<i>p</i> -value
End of data collection	89.7	83.6	-3.83	< .001

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/25 Beginning Postsecondary Students (BPS:20/25) Field Test.

Recommendations for the full-scale study

Removing telephone reminders from sample members receiving the default data collection protocol did not significantly alter respondent sample representativeness by the end of data collection. Further, while offering telephone reminders did bring in responses a few days sooner than not offering telephone reminders, this small gain in efficiency did not justify the cost of the reminders. However, the response rate of sample members receiving the default data collection protocol was higher when they were given telephone reminders to complete the survey. These results therefore indicate that telephone reminders, despite their cost, are still useful for encouraging response from sample members who have responded to NPSAS and BPS survey requests in the past. To balance the utility and cost of telephone reminders, in the BPS:20/25 full-scale data study we recommend using telephone reminders only for certain sample members in the default data collection group; specifically, sample members in the default data collection group that belong to demographic groups with low response rates. Encouraging response from these groups with telephone reminders would reduce their potential for nonresponse bias.