

2020/22 BEGINNING POSTSECONDARY STUDENTS
LONGITUDINAL STUDY
(BPS:20/22)

Full-scale

Appendix D

Results of the BPS:20/22 Field Test Experiments

OMB # 1850-0631 v.19

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

December 2021

Appendix D
Results of the BPS:20/22 Field Test Experiments

The BPS:20/22 field test included two sets of experiments: data collection experiments focused on survey participation to reduce nonresponse error and the potential for nonresponse bias (Section D.1) and questionnaire design experiments focused on minimizing measurement error to improve data quality (Section D.2). Full details of the experiments were described and approved in BPS:20/22 Field Test Supporting Statement Part B (OMB# 1850-0631 v.18).

D.1 Evaluation of Data Collection Experiments

Decreasing response rates have long been a threat to survey data quality (e.g., Massey and Tourangeau 2012) since lower response rates can increase the potential for nonresponse bias, increase survey costs, and decrease sample sizes. Two data collection experiments were designed for the BPS:20/22 field test to investigate the effects of: a) an “early bird” incentive experiment where respondents receive an additional \$5 incentive if they complete the survey within the first three weeks of data collection, and b) a survey reminder mode experiment where text message reminders exclusively used for a limited period of time during data collection instead of using telephone call reminders.

Three indicators were identified to test the effectiveness of these experiments: 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 treatments significantly increased survey response.

Next, using administrative frame data, sample representativeness was assessed across age, sex, ethnicity, race, and institutional control (i.e., public institutions; private nonprofit institutions; and private for-profit institutions). Similar variables were used to evaluate sample composition in previous NCES studies (e.g., B&B:16/20). Estimates for these characteristics were compared across respondents and nonrespondents within each of the experimental and control groups (i.e., do the experimental manipulations encourage survey participation from different kinds of sample members than the control condition?). Two-sided t-tests¹ were used assess continuous respondent characteristics (i.e., age), and Pearson chi-squared tests to assess categorical respondent characteristics (i.e., sex, ethnicity, race, and institutional control).

Finally, data collection efficiency is operationalized as the number of the days between the start of the experiment and survey completion. One-sided t-tests¹ were used to explore whether the experimental treatments significantly reduce the number of days from the start of the experiment it takes respondents to complete the survey. Table D.1 summarizes these indicators, their operationalization, and the analytic approaches.

Table D.1. 2020/22 Beginning Postsecondary Students Field Test: 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)	χ^2 test descriptive t-test - continuous variables
Sample representativeness	Compare estimates for respondents to estimates of nonrespondents	χ^2 test – categorical variables

¹ Results use Satterthwaite (1946) approximation in difference-of-means tests with unequal variances.

Data collection efficiency	Number of days from start of experiment to survey completion	t-test
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.		

The overall response rate for the BPS:20/22 field test sample² (n = 3,520) is 63.7 percent (n = 2,240). This total includes respondents who completed the full interview (n = 2,150) and those who completed part of the interview (i.e., partial completions; n = 90).

a) *Experiment #1, Data Collection: Early Bird Incentives*

Offering time-limited incentives for early responders (i.e., “early bird” incentives) can lead to faster responses and increased participation rates within a specified incentive period (e.g., LeClere et al. 2012; Ward et al. 2014; Coopersmith et al. 2016). To investigate the effects of an “early bird” incentive, the BPS:20/22 field test sample members (n = 3,520²) were randomly assigned to one of two groups: a control group (n = 1,760) who received the baseline incentive, or a treatment group (n = 1,760) who received a \$5 early bird incentive offer in addition to the baseline incentive if they completed the survey within the first three weeks of data collection.

Results.

Survey response. Response rates were observed at two time points during data collection: at the end of the three-week early bird period (i.e., 3 weeks into data collection) and at the end of data collection (17 weeks in the field). **Table D.2** summarizes the response rates for the control and early bird groups at these two time points.

At the end of the three-week early bird period, the response rate for the experimental early bird group (42.6 percent) was significantly higher than the control group (37.2 percent; $\chi^2 = 10.53, p < .01$). However, at the end of data collection, response rates for the control group (63.4 percent) and the experimental group (63.9 percent) did not significantly differ ($\chi^2 = 0.08, p = .78$). Notably, these results diverge from past research that finds increased survey participation with early bird incentives (e.g., LeClere et al. 2012; Coopersmith et al. 2016). This difference may have occurred because these past studies offered larger early bird incentives (\$20 or more) than the BPS:20/22 field test (\$5).

The results from this study therefore indicate that a \$5 early bird incentive may be effective at increasing response at the beginning of data collection, but these differences disappear as time in the field increases. This limits the utility of an early bird incentive for surveys with a longer field period, like the BPS:20/22 full-scale survey.

Table D.2. 2020/22 Beginning Postsecondary Students Field Test response rates by early bird experimental condition and evaluation period: 2021

	Control Group Baseline Incentive	Treatment Group Early Bird Plus Baseline	χ^2	p-value
End of early bird period	37.2	42.6	10.53	< 0.01
End of data collection	63.4	63.9	0.08	0.78

NOTE: Results exclude ineligible cases.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

² Excluding ineligible sample members.

Representativeness. Sample representativeness of respondents was investigated at the end of data collection. Table D.3 below displays sample composition for each of respondent characteristic by experimental condition.

BPS:20/22 field test respondents were significantly younger than nonrespondents in both the experimental early bird group ($t(981.2) = 7.55, p < 0.001$) and the control group ($t(1,012.0) = 6.92, p < 0.001$). This indicates that younger respondents were similarly overrepresented in both the experimental and control groups; the experimental group did not better represent the overall sample in terms of age than the control group. There was not, however, a significant difference between the sex of respondents and nonrespondents in both the experimental early bird group ($\chi^2 = 0.32, p = 0.57$) and the control group ($\chi^2 = 0.57, p = 0.45$). These results indicate that the percentage of female respondents in both groups was similar to the overall sample.

White respondents were overrepresented in both the early bird group ($\chi^2 = 5.24, p < .05$) and the control group ($\chi^2 = 9.98, p < 0.01$), but there was not a significant difference between percentage of Hispanic or Latino respondents and nonrespondents in the early bird group ($\chi^2 = 0.31, p = 0.58$) or the control group ($\chi^2 = 0.54, p = 0.46$). This again indicates that the experimental group did not better represent the overall sample in terms of race or ethnicity than the control group.

There was not a significant difference between the percentage of respondents and nonrespondents from public institutions in either the early bird group ($\chi^2 = 0.09, p = 0.78$) or the control group ($\chi^2 = 0.67, p = 0.41$). Respondents in both groups were therefore representative of the overall sample. Respondents from private nonprofit institutions were overrepresented in both the early bird group ($\chi^2 = 25.93, p < .001$) and the control group ($\chi^2 = 27.54, p < 0.001$). In turn, respondents from private for-profit institutions were underrepresented in the early bird group ($\chi^2 = 37.29, p < .001$) and the control group ($\chi^2 = 23.53, p < 0.001$).

Taken together, these results indicate that offering an early bird incentive encouraged response from the same types of sample members as the baseline incentive alone.

Table D.3. 2020/22 Beginning Postsecondary Students Field Test sample composition by early bird experimental condition: 2021

	Control Group Baseline Incentive	Treatment Group Early Bird Plus Baseline
Age (mean)		
Respondents	21.6	21.5
Nonrespondents	23.4	23.6
Respondents – Nonrespondents	-1.8***	-2.1***
Overall Sample ($n = 3,510$) ¹	22.3	22.2
Female (in percent)		
Respondents	58.0	57.4
Nonrespondents	56.1	56.0
Respondents – Nonrespondents	1.9	1.4
Overall Sample ($n = 3,360$) ¹	57.3	56.9
White² (in percent)		
Respondents	70.2	71.2
Nonrespondents	62.3	65.6
Respondents – Nonrespondents	7.9**	5.6*
Overall Sample ($n = 3,180$) ¹	67.6	69.4
Hispanic or Latino (in percent)		
Respondents	21.7	21.9
Nonrespondents	23.3	23.1
Respondents – Nonrespondents	-1.6	-1.2
Overall Sample ($n = 3,190$) ¹	22.2	22.2
Institution Control (in percent)		
Public		
Respondents	57.9	59.7
Nonrespondents	59.9	59.0
Respondents – Nonrespondents	-2.0	0.7
Overall Sample ($n = 3,520$)	58.6	59.5
Private nonprofit		
Respondents	28.0	27.3
Nonrespondents	17.0	16.6
Respondents – Nonrespondents	11.0***	10.7***
Overall Sample ($n = 3,520$)	24.0	23.4
Private for-profit		
Respondents	14.1	13.0
Nonrespondents	23.2	24.4
Respondents – Nonrespondents	-9.1***	-11.4***
Overall Sample ($n = 3,520$)	17.4	17.1

* $p < .05$, ** $p < .01$, *** $p < .001$

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

² "White" includes those who are and are not of Hispanic or Latino background. Hispanic or Latino is considered an ethnicity rather than a race. People of Hispanic or Latino origin may be of any race.

NOTE: Results exclude ineligible cases.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Efficiency. Data collection efficiency was examined (i.e., the average number of days it took respondents to complete the survey) at two time points during data collection: the end of the three-week early bird period (3 weeks into data collection), and 2) at the end of data collection (17 weeks in the field). Table D.4 summarizes the average number of days to survey completion for the control and early bird groups at these two time points.

At the end of the early bird period, the average number of days that it took respondents in the early bird group to complete the survey (10.2 days) was not significantly lower than the control group (8.8 days; $t(1,400.7) = -4.02$, $p = 1.00$). At the end of data collection, respondents in the experimental group took significantly fewer days (28.1 days) than respondents in the control group (30.9 days) to complete the survey ($t(2,231.7) = 2.09$, $p < 0.05$). However, this difference is small (2.8 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 D.4. 2020/22 Beginning Postsecondary Students Field Test average number of days to complete by early bird experimental condition and evaluation period: 2021

	Control Group Baseline Incentive	Treatment Group Early Bird Plus Baseline	<i>t</i>	<i>p</i> -value
End of early bird period	8.8	10.2	-4.02	1.00
End of data collection	30.9	28.1	2.09	0.02

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Recommendations for the full-scale study

Offering sample members an early bird incentive did not significantly improve response rates or representativeness by the end of data collection. Further, while offering an early bird incentive did bring in responses a few days sooner than offering the baseline incentive alone, this small gain in efficiency did not justify the cost of the early bird. Therefore, the use of an early bird incentive in the BPS:20/22 full-scale data collection is not recommended.

b) Experiment #2, Data Collection: Reminder Modes

Text message advance notifications and reminders have been shown to significantly increase response rates (e.g., Callegaro et al. 2011; Schober et al. 2015). Further, National Postsecondary Student Aid Study (NPSAS) focus group (n = 50 overall; including n = 20 FTB) results suggest that text messaging is the preferred mode of communication for most focus group participants (21 mentions). E-mails were mentioned as the preferred mode 11 times, and telephone 6 times. Text messaging was also used in NPSAS:20 to prompt sample members to complete the survey. When compared to prompting using telephone calls, these text message reminders have comparable rates of survey completes, a higher absolute number of completes, and are more cost efficient.

After two months of data collection, all nonresponding BPS:20/22 field test sample members (n = 1,840³) were randomly assigned to one of two groups: a control group that received only telephone call reminder prompts (n = 930), and an experimental group that received only text message reminder prompts (n = 910). Sample members received reminders in their assigned mode for three weeks. During this three-week reminder period, all other data collection activities (e.g., reminder e-mails, hardcopy mailings) continued for both groups.

Results.

Survey response. Response rates were examined at two time points during data collection: 1) at the end of the three-week reminder period (10 weeks in the field), and 2) at the end of data collection (17 weeks in the field). Table D.5 summarizes the response rates for the telephone reminder control group and the text message reminder experimental group at these two time points.

At the end of the reminder period, the response rate for the experimental text message group (13.6 percent) did not significantly differ from the control group who received only telephone reminders (14.2 percent; $\chi^2 = 0.12$, $p = 0.73$). Similarly, at the end of data collection, response rates for the control group (31.5 percent) and the

³ Excluding ineligible sample members.

experimental group (29.6 percent) did not significantly differ ($\chi^2 = 0.83, p = 0.36$). These results indicate that the less expensive text messaging method of sending reminders was just as effective as telephone reminders at prompting survey response.

Table D.5. 2020/22 Beginning Postsecondary Students Field Test response rates by reminder mode experimental condition and evaluation period: 2021

	Control Group Telephone Reminders	Treatment Group Text Reminders	χ^2	p-value
End of reminder period	14.2	13.6	0.12	0.73
End of data collection	31.5	29.6	0.83	0.36

NOTE: Results exclude ineligible cases.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Representativeness. Respondent sample representativeness was investigated at the end of data collection. Table D.6 below displays sample composition for each of the respondent characteristics by experimental condition.

BPS:20/22 field test respondents were significantly younger than nonrespondents in both the experimental text message reminder group ($t(602.6) = 3.87, p < 0.001$) and the control telephone reminder group ($t(813.5) = 6.72, p < 0.001$). This indicates that younger respondents were similarly overrepresented in both the experimental and control groups. There was not, however, a significant difference between the sex of respondents and nonrespondents in both the text message group ($\chi^2 = 0.91, p = 0.34$) and the telephone group ($\chi^2 = 0.01, p = 0.92$). These results indicate that the percentage of female respondents in both groups was similar to the overall sample.

White respondents were overrepresented in the control group ($\chi^2 = 7.40, p < 0.01$). However, there was not a significant difference between the percentage of White respondents and nonrespondents in the text message group ($\chi^2 = 0.28, p = 0.60$). This indicates that the text message reminders were more effective than telephone reminders at encouraging participation from respondents of races other than White. There was not, however, a significant difference between percentage of Hispanic or Latino respondents and nonrespondents in the text group ($\chi^2 = 0.87, p = 0.35$) or the telephone group ($\chi^2 = 1.92, p = 0.17$); the percentage of Hispanic or Latino respondents in both groups was similar to the overall sample.

There was not a significant difference between the percentage of respondents and nonrespondents from public institutions in either the text group ($\chi^2 = 1.53, p = 0.22$) or the telephone group ($\chi^2 = 0.16, p = 0.69$). Respondents in both groups were representative of the overall sample. Respondents from private nonprofit institutions were overrepresented in both the text message reminder group ($\chi^2 = 4.05, p < .05$) and the telephone reminder group ($\chi^2 = 13.18, p < 0.001$). Respondents from private for-profit institutions were underrepresented in the text message group ($\chi^2 = 12.03, p < .01$) and the telephone group ($\chi^2 = 9.05, p < 0.01$).

Taken together, these results indicate that generally, text message reminders encouraged response from the same types of sample members as telephone reminders. However, text message reminders have the added benefit of recruiting more respondents of races other than White than telephone reminders, making the resultant respondent sample more like the overall sample.

Table D.6. 2020/22 Beginning Postsecondary Students Field Test sample composition by reminder mode experimental condition: 2021

	Control Group Telephone Reminders	Treatment Group Text Reminders
Age (mean)		
Respondents	21.4	21.8
Nonrespondents	23.8	23.2
Respondents – Nonrespondents	-2.4***	-1.4***
Overall Sample (n = 1,830) ¹	23.0	22.8
Female (in percent)		
Respondents	55.2	53.0
Nonrespondents	55.5	56.5
Respondents – Nonrespondents	-0.3	-3.5
Overall Sample (n = 1,690) ¹	55.4	55.4
White² (in percent)		
Respondents	74.1	65.3
Nonrespondents	64.5	63.3
Respondents – Nonrespondents	9.6**	2.0
Overall Sample (n = 1,570) ¹	67.8	64.0
Hispanic or Latino (in percent)		
Respondents	21.3	23.6
Nonrespondents	25.8	20.6
Respondents – Nonrespondents	-4.5	3.0
Overall Sample (n = 1,570) ¹	24.3	21.6
Institution Control (in percent)		
Public		
Respondents	58.6	63.3
Nonrespondents	59.9	58.9
Respondents – Nonrespondents	-1.3	4.4
Overall Sample (n = 1,840)	59.5	60.2
Private nonprofit		
Respondents	26.0	23.3
Nonrespondents	15.9	17.6
Respondents – Nonrespondents	10.1***	5.7*
Overall Sample (n = 1,840)	19.1	19.3
Private for-profit		
Respondents	15.4	13.3
Nonrespondents	24.1	23.5
Respondents – Nonrespondents	-8.7**	-10.2**
Overall Sample (n = 1,840)	21.4	20.5

* $p < .05$, ** $p < .01$, *** $p < .001$

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

² "White" includes those who are and are not of Hispanic or Latino background. Hispanic or Latino is considered an ethnicity rather than a race. People of Hispanic or Latino origin may be of any race.

NOTE: Results exclude ineligible cases.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Efficiency. Data collection efficiency was examined (i.e., the average number of days it took respondents to complete the survey) at two time points during data collection: 1) at the end of the three-week reminder period (10 weeks in the field), and 2) at the end of data collection (17 weeks in the field). Table D.7 summarizes the average number of days to survey completion for the control and early bird groups at these two time points.

At the end of the three-week reminder period, respondents who received text message reminders completed the survey in 48.4 days on average, which was 4.3 days sooner than respondents who received telephone reminders (52.7 days; $t(245.5) = 1.99$, $p < .05$). However, at the end of data collection, this difference disappeared. The number of days it took for respondents in the text message reminder group to complete (75.5 days) was not significantly different from the telephone reminder group (77.0 days; $t(542.8) = 0.62$, $p = 0.27$). These results demonstrate that text

message reminders are just as efficient as telephone reminders across the entire field period, and more efficient early in data collection.

Table D.7. 2020/22 Beginning Postsecondary Students Field Test average number of days to complete by reminder mode experimental condition and evaluation period: 2021

	Control Group Telephone Reminders	Treatment Group Text Reminders	<i>t</i>	<i>p</i> -value
End of reminder period	52.7	48.4	1.99	0.02
End of data collection	77.0	75.5	0.62	0.27

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Recommendations for the full-scale study

Text message reminders achieved response rates, representativeness, and efficiency that was comparable to more expensive telephone reminders. Based on these results, the use of text reminders as a part of the BPS:20/22 full-scale data collection is recommended. However, there are still situations where telephone reminders will be useful (e.g., sample members that do not have a cell phone number, those who opt out of the text message reminders). Therefore, a mix of telephone and text message reminders is recommended for the BPS:20/22 full-scale data collection.

D.2 Evaluation of Questionnaire Design Experiments

Two field test survey instrument experiments tested different methods of collecting month-level enrollment intensity and address information to identify which method could reduce burden and result in higher quality data. In addition, two randomly assigned modules that collected information about the impacts of the coronavirus pandemic were fielded. The results of this testing informed which survey questions are best suited for inclusion in the full-scale study. Table D.8 summarizes the indicators, operationalizations, and analytic approaches used to assess results of the two experiments and comparison of the coronavirus pandemic modules.

Table D.8. 2020/22 Beginning Postsecondary Students Field Test: Overview of indicators, operationalization, and analytic approaches to survey questionnaire design experiments

Indicator	Operationalization	Analytic Approach ¹
Missingness	Compare item- and question-level nonresponse rate	<i>t</i> -test
Administrative data concordance	Agreement rates between self-reported enrollment and National Student Clearinghouse (NSC) administrative records	<i>t</i> -test
Timing burden	Mean timing burden at the survey question-level	<i>t</i> -test
Response patterns	Mean timing burden of entire coronavirus pandemic module	descriptive
	Rate of contradictory response about enrollment at primary institution	
	Rate of straightlining on largest grid-format question	
		<i>t</i> -test

¹ All analyses compare results by experimental group assignment, or by coronavirus pandemic module assignment when applicable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

a) Experiment #3, Survey Instrument: Collection of Month-level Enrollment Intensity

Enrollment intensity is an important measure for BPS:20/22 due to its role in financial aid eligibility and other analyses of student persistence and credential attainment. The BPS:20/22 field test survey randomly assigned respondents into one of two methods of collecting enrollment information to determine which method increases the reliability of self-reported enrollment information. The first method, comparable to past BPS

surveys (control group), used a forced-choice grid question that collected enrollment intensity for each academic year (B22ASTST2) on a single form. The treatment group received the second method, in which respondents were administered separate questions for full-time and part-time enrollment intensity for specified academic years (B22ANENRFLT, B22ANENRLPT). Additionally, this experiment was conducted to analyze which method may potentially reduce respondent burden, as the burden associated with navigating the custom calendar survey questions is high. Following the control or experimental gate(s), both groups are administered the custom calendar survey questions, which display individual month buttons by academic year, and respondents may choose a “select all” option to select all months in the academic year or select months individually. This burden concern is particularly important given the BPS:20/22 full-scale survey will collect information about two years of academic enrollment (24 months), rather than one academic year (12 months) as in the field test. Following the administration of each version, specific months of enrollment were collected for the applicable enrollment intensity. The following analysis is limited to results from respondents reporting continued enrollment at their NPSAS institution.

Results.

Of the 2,240 BPS:20/22 field test respondents, 1,320 respondents are included in this analysis (59 percent). Table D.9 provides the randomized experimental assignment for field test respondents included in the analysis, and cases excluded from the analysis.

Table D.9. 2020/22 Beginning Postsecondary Students Field Test Month-level Enrollment Intensity Analysis Cases: 2021

Respondents	Number of cases	Percent
Total	2240	100
Analysis Cases	1320	58.9
Control: single forced-choice grid	680	30.4
Treatment: yes/no radio gates by intensity	640	28.5
Excluded from analysis¹	920	41.1
No NPSAS enrollment after base-year	790	35.1
NPSAS not NSC reporting institution	90	3.8
Enrollment intensity timing outliers	50	2.1

¹ Field test respondents were excluded if they indicated not continuing enrollment at NPSAS and not administered enrollment intensity questions for their NPSAS base-year degree, respondents whose NPSAS institution does not report to NSC, and respondents identified as timing outliers for the NPSAS enrollment intensity gate(s). Outliers were calculated by normalizing the data and excluding extreme values identified by Tukey’s formula (1977) which is not sensitive to distributional assumptions. Respondents with timing burden for the enrollment intensity gate(s) outside of the interquartile range (IQR = 3rd quartile - 1st quartile) multiplied by 1.5 were identified as timing outliers.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Missingness. None of the respondents assigned to the control group left the enrollment intensity gate missing (0 percent), and 0.3 percent of respondents left either of the yes/no radio gates missing in the treatment group, this difference in missingness is not statistically significant ($t(636) = 1.42, p = 0.1575$)⁴. Table D.10 displays month-level enrollment intensity missingness by experimental group assignment.

⁴ Results use Satterthwaite (1946) approximation in difference-of-means tests with unequal variances.

Table D.10. 2020/22 Beginning Postsecondary Students Field Test Month-level Enrollment Intensity Missingness: 2021

Group	Number of cases	Percent missing
Control: single forced-choice grid	680	0
Treatment: yes/no radio gates by intensity	640	0.3

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Concordance with administrative data source. To determine if the enrollment intensity gate question(s) impact the accuracy of enrollment data collected, field test respondents' NPSAS base-year degree enrollment reported in the survey data were compared with administrative enrollment data from the National Student Clearinghouse (NSC). Enrollment was compared for the months of July 2020 through December 2020, a retrospective time period before BPS:20/22 field test data collection. The survey response was considered to be in concordance with administrative data from NSC if the respondent either a) reported no enrollment for that intensity during the six months of interest, and there are no enrollment records in NSC for NPSAS institution during that same timeframe; or, b) reported enrollment for that intensity during the six months of interest, and NSC has a record of the respondent attending NPSAS for that same intensity at any point in those same six months of interest.

Differences in concordance between the treatment and control group were not statistically significant. For both experimental groups, 70 percent of respondents were in concordance with NSC enrollment records during the timeframe of interest regardless of intensity ($t(1311.1) = 0.24, p = 0.8119$). Full-time enrollment had the highest concordance, the control group had an 89 percent concordance rate, and the treatment group an 87 percent concordance rate ($t(1292.8) = 1.04, p = 0.2991$). See [Table D.11](#) for concordance rates overall, and by enrollment intensity.

Table D.11. 2020/22 Beginning Postsecondary Students Field Test Month-level Enrollment Intensity Concordance with NSC, overall and by intensity: 2021

Group	Number of cases	Concordance rates		
		Overall ¹	Full-time	Part-time
Overall agreement	1320	70.0	88.1	77.9
Control: single forced-choice grid	680	70.0	89.0	77.6
Treatment: yes/no radio gates by intensity	640	70.6	87.1	78.4

¹ Overall concordance between field test response and NSC administrative records was calculated irrespective of enrollment intensity (e.g., if NSC records indicated full-time enrollment between the same months, and respondent indicated part-time enrollment between the same months, it would be considered an agreement between the sources). Therefore, totaling full-time and part-time agreement rates will not sum to total of overall agreement rate.

NOTE: Concordance rates were determined by comparing the same six-month timeframe between field test response and NSC enrollment records, for which enrollment could be determined (e.g., enrollment that had occurred prior to the field test survey data collection period), from July 2020 to December 2020. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Timing burden. Across both treatment and control groups, answering the enrollment intensity gate(s) took 13.7 seconds, on average (see [Table D.12](#)). The control group

took an average of 10.5 seconds to answer the single force-choice grid question. The treatment group took significantly longer, at an average of 17.2 seconds to answer both of the separate yes/no radio gates by enrollment intensity ($t(1312.8) = 15.47$, $p < .0001$). This difference in time is expected given the additional screen respondents must navigate in the treatment group. An important caveat with timing burden is that field test respondents were only asked to report on enrollment for one academic year. For the full-scale survey, respondents will be asked to report their enrollment for two academic years, which will increase the size of the forced-choice grid (e.g., the grid will have two years, instead of just one), though the anticipated burden increase of the forced-choice grid is unlikely to exceed that of the separate yes/no radio gates by enrollment intensity.

Table D.12. 2020/22 Beginning Postsecondary Students Field Test Month-level Enrollment Intensity Timing Burden, overall and by treatment group: 2021

Group	Number of cases	Mean time (in seconds)
Overall timing burden	1320	13.7
Control: single forced-choice grid	680	10.5
Treatment: yes/no radio gates by intensity ¹	640	17.2

¹ Timing burden for the full-time and part-time enrollment questions were combined to calculate total burden for respondents in the treatment group.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test.

Recommendations for the full-scale study

Overall, the method of collecting enrollment intensity does not impact missingness rates or reporting accuracy (as determined by administrative concordance). However, administering the separate intensity yes/no radio questions does significantly increase respondent burden by an average of 6.7 seconds (64 percent increase), though the timing difference may be reduced in the full-scale survey when the grid form includes two academic years of enrollment instead of one. Additionally, cognitive testing conducted in Spring 2020 assessed these two approaches of collecting enrollment intensity (BPS:20/22 Field Test Appendix D: Cognitive and Usability Testing Summary (OMB# 1850-0631 v.18)). A majority of participants preferred the forced-choice grid. These participants generally reported that the visual layout of this item made it seem more concise and allowed them to think about and choose their answer more easily. Given the field test and cognitive interview results, data will continue to be collected using the single forced-choice grid in the BPS:20/22 full-scale survey instrument.

b) Experiment #4, Survey Instrument: Collection of Address Information Using Predictive Search Database

Predictive search forms have become commonplace in web-based data collections. The collection of respondent addresses is critical for survey incentive payments and future locating efforts. Therefore, the BPS:20/22 field test implemented a predictive search method of obtaining address information to increase data quality, and lower respondent burden typically associated with manual entries. The BPS:20/22 field test collected addresses for the entire responding field test sample using a predictive search format, with each address entry as a single textbox linked to an underlying

Experian database of U.S. addresses standardized to USPS guidelines. The address textbox entry allowed results of matching addresses to populate while the respondent entered an address. The respondent then selected the best match from the list of results provided, or if no matching result found, the respondent could manually enter in each address field. This analysis compares the BPS:20/22 field test respondents' predictive search addresses collection in regard to timing and missingness with the traditional collection of this information collected from the same set of BPS:20/22 field test respondents in the NPSAS:20 full-scale survey to determine the feasibility, and benefits of administering the predictive search address coder in the full-scale survey.

Results. Of the 2,240 BPS:20/22 field test respondents, 1,620 respondents are included in this analysis (72 percent). Table D.13 provides information about the cases included and excluded from the analysis. To be included in the analysis, a case must have been a respondent in both NPSAS:20 full-scale and BPS:20/22 field test, and not be in an exclusion category from either study. The results described below are from the first address collection form in the survey, *respondent permanent address* (N20G1ADR, B22G1ADR), which was administered to all 1,620 respondents included in the analysis.

Table D.13. 2020/22 Beginning Postsecondary Students Field Test Experian Address Analysis Cases: 2021

	Number of cases	Percent
Total	2240	100
Analysis Cases	1620	72.3
Excluded from analysis¹	620	27.7
NPSAS:20 nonrespondent	340	15.1
Permanent address timing outlier ²	250	10.9
NPSAS:20 completed in Spanish	20	0.7
Partial interview	10	0.6
Coded foreign permanent address	10	0.3

1 Cases could be excluded for multiple reasons; therefore, the exclusion categorization was prioritized according to the order listed in the table. For example, if a NPSAS:20 nonrespondent coded a foreign permanent address in BPS:20/22 field test, the case was categorized as having been excluded due to their NPSAS:20 nonresponse status.

2 Respondents identified as being time outliers for the permanent address collection from either study (N20G1ADR, B22G1ADR), or respondents for which the timing data was unavailable. Outliers were calculated by normalizing the data and excluding extreme values identified by Tukey's formula (1977) which is not sensitive to distributional assumptions. Respondents with timing burden for the enrollment intensity gate(s) outside of the interquartile range (IQR = 3rd quartile - 1st quartile) multiplied by 1.5 were identified as timing outliers.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test, and 2019-20 National Postsecondary Student Aid Study (NPSAS:20).

Missingness. Overall, 98 percent of respondents provided a complete permanent address in NPSAS:20, compared to 85 percent in BPS:20/22 field test ($t(2023.5) = 13.88, p < .0001$)⁵. NPSAS:20 consistently yielded a higher rate of complete addresses, across all modes of administration (see Table D.14), and when controlling for mode change across the studies. Telephone mode had the highest completion rate for both studies.

Table D.14. Permanent address completion rate by study, overall and by mode of administration: 2019-2021

Study	N	Mode of administration						
		Overall Percent	Web nonmobile		Web mobile		Telephone	
			N	Percent	N	Percent	N	Percent
NPSAS:20 full-scale	1620	98.4	800	98.5	780	98.2	50	100
BPS:20/22 field test	1620	85.5	830	89.1	740	81.1	50	91.7

NOTE: Detail may not sum to totals because of rounding.

⁵ Results use Satterthwaite (1946) approximation in difference-of-means tests with unequal variances.

Timing burden. Overall, it took respondents an average of 29.5 seconds to respond to the permanent address form in NPSAS:20 full-scale, significantly higher than the 27.1 seconds it took respondents in BPS:20/22 field test ($t(3055.7) = 4.09, p < .0001$). Table D.15 provides results of timing burden overall, and by mode of administration across the studies.

The use of the Experian coder version of the permanent address collection in BPS:20/22 field test was faster than NPSAS:20 full-scale in all modes of administration, except for web nonmobile mode (though the difference is not statistically significant). Respondents who completed by web mobile mode, on average, spent 5 seconds less to provide an address in BPS:20/22 than respondents who completed by web mobile mode in NPSAS:20 ($t(1500.8) = 6.09, p < .0001$). The most impactful burden-saving was for those who completed over the telephone, interviewers spent 18 seconds longer, on average, administering the permanent address in NPSAS:20 full-scale, compared to BPS:20/22 field test ($t(86.7) = 4.68, p < .0001$).

The above analysis focuses on the permanent address form which all respondents are administered. It is important to note the cumulative reduction of burden resulting from the Experian coder is dependent on number of addresses collected. Respondents may be administered up to six address forms to collect information for parents and guardians and another friend or family member for future study contacting purposes.

Table D.15. Permanent address timing burden by study, overall and by mode of administration: 2019-2021

Study	Mode of administration							
	Overall		Web nonmobile		Web mobile		Telephone	
	N	Mean time (in seconds)	N	Mean time (in seconds)	N	Mean time (in seconds)	N	Mean time (in seconds)
NPSAS:20 full-scale	1620	29.5	800	23.4	780	33.4	50	72.3
BPS:20/22 field test	1620	27.1	830	24.2	740	28.5	50	54.1

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2020/22 Beginning Postsecondary Students (BPS:20/22) Field Test, and 2019-20 National Postsecondary Student Aid Study (NPSAS:20).

Recommendations for the full-scale study

As a result of implementing the Experian coder to collect permanent address, collection of this address required less time for BPS:20/22 field test survey respondents compared to NPSAS:20 full-scale survey respondents. However, NPSAS:20 full-scale obtained a higher rate of complete addresses (i.e., all elements are provided, including street address, city, state, and zip code). While more complete addresses were obtained without the use of the Experian database in NPSAS:20, this does not indicate all entries were error free and valid. However, by definition, all addresses obtained from the BPS:20/22 field test Experian database are complete and accurate (i.e., a true verifiable address), reducing the labor costs associated with data collection staff resolving “undeliverable” check addresses (i.e., check sent back as not a real address). For NPSAS:20, around 1.2 percent of FTB check addresses were considered “undeliverable”, and BPS:20/22 field test did not obtain any “undeliverable” addresses. Survey instrument functionalities can improve the completion rate for BPS:20/22 full-scale by implementing conversion text and soft check validations when address fields are left incomplete. These functionalities were not incorporated in the BPS:20/22 field test in order to analyze initial respondent behavior with this new coder, and survey instrumentation results from past studies indicate these are

effective methods of improving nonresponse or partial response. Given the implementation of improved instrument functionality to reduce incomplete address entries, higher quality data resulting from Experian verifiable addresses, and the significant reduction in burden using Experian compared to manual entry, the continuation of Experian database is proposed for BPS:20/22 full-scale.

D.3 Evaluation of Coronavirus Pandemic Items

The COVID-19 pandemic has created unprecedented disruptions for postsecondary education. BPS:20/22 full-scale sample members were first time beginning students during the 2019-20 academic year, when the pandemic began. Researchers, including members of the BPS:20/22 Technical Review Panel (TRP), have expressed interest in using BPS:20/22 data to examine impacts of COVID-19 on postsecondary students. To maximize opportunities to test of questions about impacts of COVID-19, BPS:20/22 field test respondents were randomly assigned into two groups that received one of two modules asking about the coronavirus pandemic. The questions in module one included revised questions from the NPSAS:20 full-scale survey (see OMB #1850-0666 v.29), and collected information about attendance, general experiences, refunds received, and institutional communication and information provided to students (questions: B22FCOVATND, B22FCOVEXPA, B22FCOVEXPB, B22FCOVRFND, B22FCOVTECH, B22FCOVCOMM, and B22FCOVINFO). Module two collected a new set of constructs identified during the BPS:20/22 field test TRP that may be of analytic value to researchers and policymakers, such as changes in enrollment and borrowing, changes in academic engagement, and access to support resources (questions: B22FCVATND2, B22FCVACAD, B22FCVNOATND, B22FCVATNDPS, B22FCVEXP2, B22FCVTHINK, B22FCVPAY, and B22FCVPERS)⁶.

Since each module collected distinct constructs, direct comparisons of results were not intended and are not presented. Rather, this randomized assignment of modules allowed a greater number of survey questions to be fielded, providing more data to inform full-scale decisions while reducing burden to individual field test respondents. Results of these questions were shared with the TRP and NCES to assist with decision-making for a final set of COVID-19 questions to be included in the BPS:20/22 full-scale survey. Questions are assessed using common measures such as response timing, item missingness, and response patterns (e.g., contradictory responses or responses lacking variation).

Results. Of the cases eligible for analysis, the respondents were equally split between the random module assignment, 50.4 percent in the module one group, and 49.7 percent in the module two group⁷.

⁶ The wording for these questions can be found in appendix E of the BPS:20/22 field test package (OMB #1850-0631 v.18).

⁷ Analyses of coronavirus pandemic questions are limited to 1,180 respondents (53 percent of all field test respondents) who: 1) completed the entire survey, and 2) reported attending their primary institution in the 2020-2021 academic year in the enrollment section of the survey. This is because most coronavirus pandemic questions were administered to respondents who attended their primary institution between July 2020 - December 2020. Additionally, of the respondents with a primary institution as defined above, cases without 3) timing data available due to completing the survey in more than one session, or 4) identified as total timing outliers were also excluded from this analysis. To detect total time outliers, the distribution of all survey times (highly right-skewed) was first normalized using a Box-Cox power transformation (Box & Cox, 1964). Cases were then removed from the overall total using an interquartile range formula adopted from Tukey (1977) with a multiplier of 1.5. Cases were excluded as outlier if total time > 75th percentile + (1.5 * interquartile range), or if total time <

Timing burden. Regardless of module assignment, the coronavirus pandemic questions took respondents an average of 2.7 minutes to complete, just under the allotted goal of 3 minutes. Respondents administered the coronavirus module by a telephone interviewer took an average of 6.2 minutes, significantly longer than web nonmobile and web mobile (2.7 minutes, $(t(50.61) = 9.43, p < .0001)$, and 2.5 minutes, $(t(51.57) = 9.84, p < .0001)$, respectively). The most burdensome questions administered were large Likert scale grids, questions that contain multiple response options within a grid with scaled response options (e.g., “Strongly agree” to “Strongly disagree”). For those respondents assigned to module one, the 6-item Likert grid *Helpful communication from primary school* (B22FCOVCOMM) took respondents an average of 42 seconds to answer. Likewise, the 7-item Likert grid for respondents assigned to module two was *Social/academic experiences at primary school during COVID-19* (B22FCVTHINK), and also took respondents an average of 42 seconds to answer.

Missingness. Across both modules, the average item nonresponse rate was 2 percent. Module one had an average nonresponse rate of 3 percent, significantly higher than the 0.6 percent nonresponse rate of module two ($t(836.72) = 5.16, p < .0001$). For module one, the question that increased overall nonresponse rate was *General experiences during COVID-19* (B22FCOVEXPB), for which 10 percent of respondents left all items on this checkbox list missing, likely due to the design of the form. If a respondent did not select any of the individual experiences, nor opted to select “None of the above”, it was considered a nonresponse. Comparatively, for module two, no single item on any of the questions administered had a higher nonresponse rate than 1 percent.

Response patterns. To analyze the response patterns of coronavirus pandemic questions, the rate of contradictory responses and straightlining (i.e., selecting the same scaled response for the entire set of response options on a grid) was examined.

To determine contradictory responses, the rate at which field test respondents disagreed regarding their primary institution attendance during the coronavirus pandemic was analyzed. Disagreement is defined as the respondent indicating they attended their primary institution between July 2020 and December 2020 in the enrollment section of the main survey but indicated not having attended primary institution in the coronavirus pandemic module during this same timeframe. Six percent of respondents across both modules provided contradictory information between the enrollment section of the survey and the coronavirus pandemic module. Respondents administered module two had a significantly higher rate of disagreement at 9 percent, compared to just 2 percent for module one ($t(857.15) = 5.35, p < .0001$). Much of this difference in disagreement is driven by design differences between the modules, module one only had one question related to their enrollment at primary institution (*Attended primary school during COVID-19*, B22FCOVATND), whereas module two had two opportunities to report enrollment information (*Attended any postsecondary institution during COVID-19*, B22FCVATND2, and *Attended primary school during COVID-19*, B22FCVATNDPS). For module two, if a respondent indicated not attending any postsecondary institution during the timeframe of interest, they were not administered the follow-up specific to their primary institution.

25th percentile - (1.5 * interquartile range).

The rate of straightlining was calculated for the largest grid-format question administered to each module group, straightlining in of itself does not necessarily indicate poor question performance as some lack of variation in response is likely valid (e.g., selecting “Strongly disagree” for all, if the respondent had negative experiences during COVID-19). However, comparison across similarly constructed questions can indicate respondent fatigue or lack of comprehension. For module 1, the response patterns of the 6-item Likert grid *Helpful communication from primary school* (B22FCOVCOMM) were analyzed, and for module two, the response patterns of the 7-item Likert grid *Social/academic experiences at primary school during COVID-19* (B22FCVTHINK) were analyzed. Across both modules, the average rate of straightlining on the largest Likert grid was 17 percent. The rate of straightlining for respondents in module one was 20 percent, significantly higher than the 12 percent for module two ($t(1036.1) = 3.43, p < .001$).

Recommendations for the full-scale study

The BPS:20/22 full-scale survey instrument will administer a subset of the questions from both field test coronavirus pandemic modules, based upon field test performance and TRP feedback. The coronavirus pandemic module for the full-scale maintains the burden goal of three minutes. In general, response timing, item missingness, and response patterns indicate that both modules performed well in the field test. The most burdensome grid with the highest straightlining observed in the field test *Helpful communication from primary school* (B22FCOVCOMM) will not be recommended for the full-scale survey. The two-step enrollment questions from module two, will be included in the full-scale survey in order to collect information about whether the coronavirus pandemic impacted a respondent’s decision to not enroll at all, or not enroll at their primary institution. The contradictory response rate was higher for this version in the field test, though the contradictory response rate is anticipated to decline as the full-scale survey will ask about enrollment for an entire academic year, rather than a single term. Additionally, adding on screen wording to remind respondents of their response from the enrollment section or implementing a validation can improve accuracy for these coronavirus pandemic gate questions (e.g., if respondent indicates attending primary institution in the enrollment section, but indicated “no” on B22FCVATND2, a soft check will alert the respondent of the conflicting answers, and provide an opportunity to correct).

References

- Box, G.E.P., and Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Statistics Society, Series B*, 26: 211-234.
- Callegaro, M. Ayhan, O., Gabler, S., Haeder, S., & Villar, A. (2011). Combining landline and mobile phone samples. A dual frame approach. Working Papers 2011/13. Gesis Leibniz-Institut fuer Sozialwissenschaften.
- Coopersmith, J., Vogel, L.K., Bruursema, T., and K. Feeney. (2016). Effects of Incentive Amount and Type of Web Survey Response Rates. *Survey Practice*, 9(1): <https://doi.org/10.29115/SP-2016-0002>.
- 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*, 4156 - 4165.
- Massey, D.S., Tourangeau, R. (2012). The Nonresponse Challenge to Surveys and Statistics New York: Sage; 2013. *Annals of The American Academy of Political and Social Science*; No. 645.
- Satterthwaite, F.E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, 2(6): 110-114.
- Schober, M.F., Conrad, F.G., Antoun, C., Ehlen, P., Fail, S., Hupp, A.L., Johnston, M., Vickers, L., Yan, Y.H., and Zhang, C. (2015). Precision and Disclosure in Text and Voice Interviews on Smartphones. *PLoS ONE* 10(6): e0128337. doi:10.1371/journal.pone.0128337
- Tukey, J.W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Ward, C., Stern, M., Vanicek, J., Black, C., Knighton, C., and L. Wilkinson. 2014. Evaluating the Effectiveness of Early Bird Incentives in a Web Survey. 2014 Federal CASIC Workshops.