

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Contents

Note.....	3
Design.....	3
Study Universe.....	5
<i>Inclusion Criteria</i>	5
<i>Exclusion Criteria</i>	5
<i>Covariates: Baseline Student Characteristics</i>	5
<i>Covariates: Baseline Parent Characteristics</i>	6
<i>Covariates: Final List of Controls</i>	6
<i>DV: Outcomes</i>	6
Randomization	6
<i>Universes</i>	6
<i>Stratification Procedure for Main Universe:</i>	7
<i>Stratification Procedure for Extreme Absence Universe:</i>	7
<i>Stratification Procedure for Low Absence Universe:</i>	7
<i>Balance Check</i>	8
Defining Treatment	8
<i>Decision Rules</i>	9
Primary Analysis.....	11
<i>Impact on Days Absent</i>	11
<i>Hypothesis Testing</i>	11
<i>Estimation</i>	12
<i>Sensitivity Checks</i>	13
<i>“Sibling” Analysis</i>	13
Exploratory Subgroup Analysis for the Primary Outcome.....	14
<i>Endogenous Stratification of Absences</i>	14
<i>High School vs. All Other Grade Levels of School</i>	14
<i>School-specific Communications</i>	14
Analysis of Secondary Outcomes	14
Missing Outcomes	14
Survey Analysis	15
<i>Inclusions</i>	15
<i>Analysis Methods</i>	15

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

<i>Internal Validity</i>	15
<i>Impact on Parental Beliefs</i>	16
<i>Impact on Parental Behavior</i>	17
Additional Analyses.....	17
Appendix.....	18

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Note

We do not have permission to identify our partner district, so its name has been redacted from this pre-analysis plan.

Design

This experiment evaluates the effectiveness of using parental engagement to improve student attendance. We address three main research questions:

RQ1: Does contacting guardians and encouraging them to improve their students' attendance reduce absences?

RQ2: Does communicating to guardians the total number of days their student missed reduce absences?

RQ3: Does communicating to guardians the total number of days their student missed *as compared to the absences of a typical student* reduce absences?

We also address the exploratory research question:

RQ4: Do these interventions also impact the attendance of other students in the household not explicitly mentioned in the mailings?

The three main research questions can be condensed into three behavioral-psychology-motivated components:

1. ENCOURAGE – Reaching out to guardians and encouraging them to improve student attendance
2. SELF – Informing guardians about their students' absences
3. NORMS – Comparing students' absences to what's "normal"

There is no way of testing the effectiveness of the NORMS component without also implicitly measuring the effectiveness of generic parental encouragement (i.e., ENCOURAGE) and giving guardians information about their students' attendance (i.e., SELF). As such, we leverage an experimental design that tests the *added impact* of each of these three components. Participating households are randomly assigned to one of four conditions:

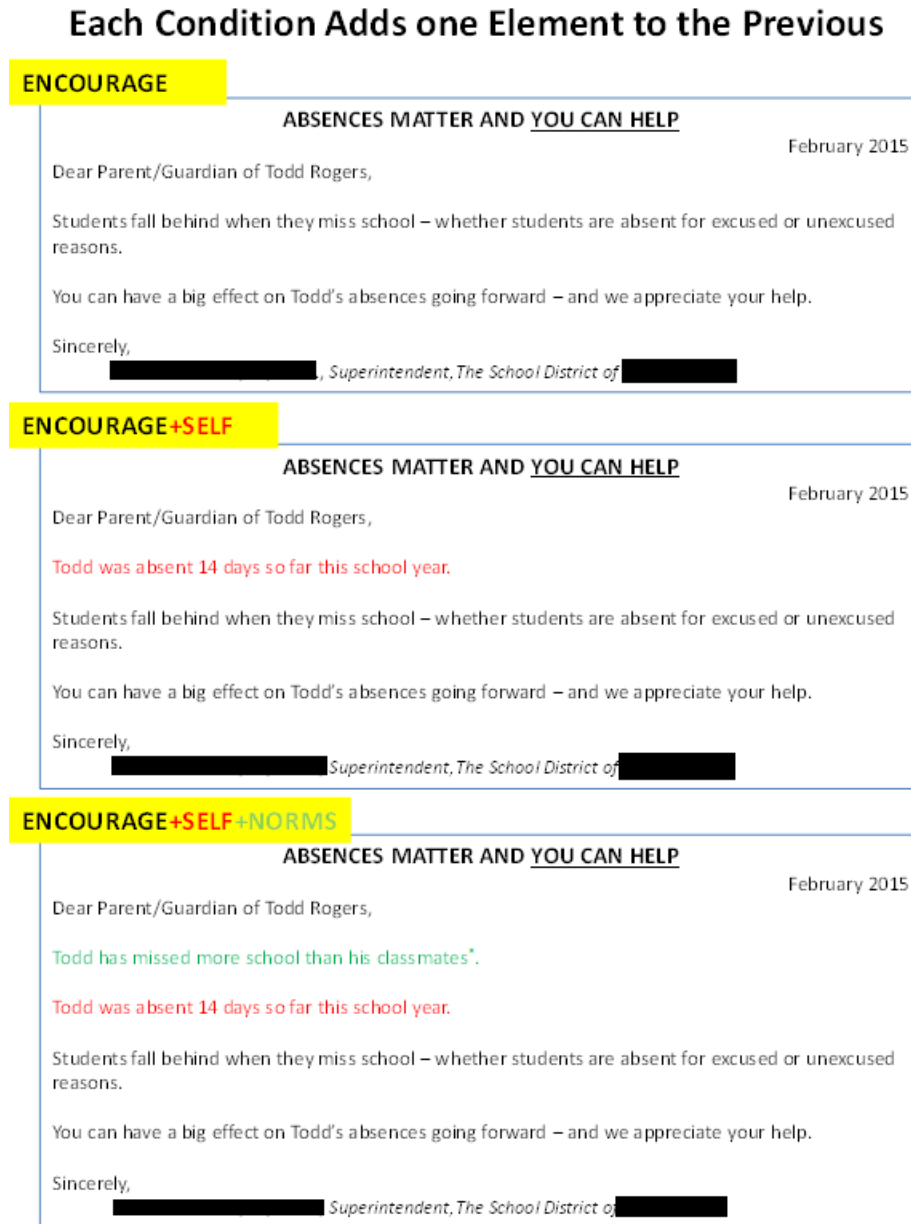
- (1) CONTROL: receives no additional outreach as part of this experiment.
- (2) ENCOURAGE: receives mail stressing the importance of attendance, parental efficacy, and absence-reduction as part of parental role.
- (3) ENCOURAGE + SELF: receives mail that has the same content as the ENCOURAGE mail but also states the number of days of school the guardians' student missed.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

- (4) **ENCOURAGE + SELF +NORMS:** receives mail that contains all the same information as the **ENCOURAGE + SELF** mail but also compares the student's attendance record to that of the typical student in their school and grade.

See Figure 1, for an illustration of this additivity.

Figure 1



Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Study Universe

All regular-status, public elementary, middle, and high schools in a large urban school district were included in the experimental universe.¹

We included only those students for whom the district had data from the 2013-2014 SY. Because the district did not have access to pre-school information, most of our experimental universe is composed of students who in the 2014-2015 SY are enrolled in grades 1-12.

Inclusion Criteria

- We only included households that were sent a consent form and did not opt out. We also excluded households whose home language was updated in the district data to something other than the languages in which the consent form was sent.
- We only target one student in each household. If multiple students at an address met the below criteria, one student from that address was randomly selected into the sample universe and the treatment messages focused on this student. We refer to this student as the *focal student*.

Exclusion Criteria

- Students who were in grade 12 as of 2012-2013
- Students who were NOT enrolled as of June 2014
- Currently Enrolled flag equals "N"
- Students flagged as graduated, duplicate, logical delete, or withdraw
- Homeless Flag equals "Y"
- Alternative Education School Flag equals "Y"
- Disability Flag equals "Y"
- Home Language was not one of our consent form languages (English, Albanian, Arabic, Mandarin, French, Khmer, Russian, Spanish, Vietnamese), blank, or undefined
- Pilot study schools
- Students who were in a pilot study school at the time of the pilot but may be in a different school now
- Opt-outs and bouncebacks
- A Virtual Academy
- Households with more than 7 eligible students

Covariates: Baseline Student Characteristics

For each student, we obtained the following baseline characteristics from administrative records:

- Sex
- School and grade
- Race (Black; Hispanic; White; Asian; Other)
- Free and Reduced Lunch Status
- LEP Status
- Number of days absent in the 2013-2013 SY
- Number of days absent in the 2014-2015 SY prior to the first treatment mailing

¹ "Regular-status" schools are defined as all public schools in the district that do not have a special designation, such as specialized schools (like schools for the blind), alternative education schools, and charter schools.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Covariates: Baseline Parent Characteristics

For each family, we obtained the following baseline characteristics from administrative records:

- Number and grade of students enrolled in another participating public school
- Home language

Covariates: Final List of Controls

The following is therefore our standard set of “control variables,” unless otherwise stated:

- Female
- English is the home language
- Free and Reduced Lunch Indicator
- LEP Indicator
- Indicator for Black/African-American race
- Number of days absent in the 2013-2014 SY
- Number of days absent in the 2014-2015 SY prior to the first treatment mailing

We will also include the following fixed effects:

- School indicator
- Grade indicator

DV: Outcomes

We will obtain the following outcomes directly from administrative records:

- **Primary outcome:** Number of days recorded as absent between the day immediately after postcard #1 was sent out and the end of school year
- **Secondary outcomes:**
 - o Number of tardies after postcard #1 was sent out through end of school year
 - o Grades of core courses averaged across the final three marking periods of the 2014-2015 SY
 - o State standardized reading and math test scores for all available grades

Randomization

Universes

We expected that the interventions would differentially affect students based on previous absences. For instance, students with very good attendance might not be affected at all due to ceiling effects. Extreme absences, on the other hand, may be the results of data errors or represent students with undocumented, singular life circumstances (e.g., sudden hospitalization, homelessness). As such, the study had three separate experimental universes:

1. Main universe (students who missed at least three days more than the minmode² of their school-grade but are not two standard deviations above their school-grade mean). The four treatments included were:
 - a. CONTROL
 - b. ENCOURAGE
 - c. ENCOURAGE + SELF
 - d. ENCOURAGE + SELF + NORMS (sample size, 1:1:1:1)

² If there are multiple modes, we select the minimum mode. We found that more individual students are clustered around the minimum mode as compared to the maximum mode, so we concluded the minmode is a better representation of the “typical” student.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

2. Extreme absence universe (students with absences greater than 2 standard deviations above their school-grade mean). The four treatments included were:
 - a. CONTROL
 - b. ENCOURAGE
 - c. ENCOURAGE + SELF
 - d. ENCOURAGE + SELF + NORMS (sample size, 1:1:1:1)
3. Low absence universe (all other students with 1 or more absences³ not already included in the above universes). The three treatments included were:
 - a. CONTROL
 - b. ENCOURAGE
 - c. ENCOURAGE + SELF (sample size, 1:1:1)

Stratification Procedure for Main Universe:

Select all students who qualify for main universe (students who missed at least three days more than the minmode of their entire school-grade but are not two standard deviations above their school-grade mean). If there is more than one eligible student in a household, one student is randomly selected for inclusion. For every school-grade at the high school level, students are broken up into attendance quartiles (least absent quartile to most absent quartile) for the purposes of stratification. For every school-grade at grades K-8, students are broken up into attendance halves (least absent and most absent). If a resulting stratum (school-grade-quartile or halve) has fewer than four⁴ qualifying students, those students' school-grade is used as their stratum (with no further distinctions based on absenteeism). If any school-grade has fewer than four qualifying students, those students are excluded from the experiment.

Stratification Procedure for Extreme Absence Universe:

Select all students who are not in the main absence universe and have absences in excess of two standard deviations of their school-grade mean. (The extreme absence cutoff points varied considerably across school-grades, so we did not use a single uniform cutoff point.) If a household already had a student assigned to the main universe, the household is excluded. If there is more than one student with extreme absences in a household, one student is randomly selected for inclusion. Due to the small sample size of the extreme absence universe, we stratify by school. If any school has fewer than four⁴ extreme absence students, those students are excluded from the experiment.

Stratification Procedure for Low Absence Universe:

Select all students who are not in the main or extreme absence universe. If a household already had a student assigned to the main or extreme absence universe, the household is excluded. We exclude all students with perfect attendance records. If there is more than one student with low absences in a household, one student is randomly selected for inclusion. For every school-grade at the high school level, students are broken up into attendance quartiles (least absent quartile to most absent quartile) for the purposes of stratification. For every school-grade at grades K-8, students are broken up into attendance halves (least absent and most absent). If a resulting

³ Students with perfect attendance in the previous year were excluded from the study.

⁴ Because there are four treatment conditions in this universe, and each stratum should have students in all treatment conditions, four is the minimum size for the stratum.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

stratum (school-grade-quartile or halve) has fewer than three⁵ qualifying students, those students' school-grade is used as their stratum (with no further distinctions based on absenteeism). If any school-grade has fewer than three qualifying students, those students are excluded from the experiment.

Balance Check

We performed balance checks and found that randomization was balanced across key covariates.

Defining Treatment

We faced two key practical constraints in administering treatment:

Students with excellent attendance

First, the district asked us to refrain from mailing Round 3, Round 4, and Round 5 treatments to students who had been absent less than 3 days in the most current attendance data of the 2014-2015 School Year. Since random assignment was conducted based on 2013-2014 absences, this adjustment simply meant that some students in each condition did not receive additional pieces of mail some of the time.

Students with better-than-expected attendance

Second, mailing rounds 2, 3, 4, and 5 reflect up-to-date absence records from the 2014-2015 academic year. While inclusion in the experiment required that students have been absent more days than their classmates in 2013-2014, in some cases, those same students were absent *fewer* days than their classmates in 2014-2015. We committed at the beginning of this experiment to not send messages reporting that students were absent *less* than their classmates, because research shows that people conform to the behavior of others. Theory and evidence suggests that such messages would increase absences (Schultz et al, 2007). Therefore, for such students assigned to the ENCOURAGE+SELF+SOCIAL condition, we sent their guardians the ENCOURAGE+SELF mailing instead of the ENCOURAGE+SELF+SOCIAL mailing. The rationale for this is based on Research Question #3:

RQ3: Does communicating to guardians the total number of days their student missed as compared to the absences of a typical student reduce absences?

RQ3 is focused on the impact of adding the social norm message to the ENCOURAGE+SELF mailing. If the social norm message is no longer applicable, guardians may as well receive the ENCOURAGE+SELF mailing so as to ensure that they continue receiving some form of treatment.

⁵ Because there are three treatment conditions in this universe, and each stratum should have students in all treatment conditions, three is the minimum size for the stratum.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Decision Rules

Table 1 illustrates the exact decision rules followed at each round of mailings to implement these changes.

Attendance Project in a Major Metropolitan School District
 2014-2015 SY
 Pre-Analysis Plan

Table 1. What absence record is needed to receive each condition's mailing, each round of mailing?

Condition assignment

	Control	ENCOURAGE	ENCOURAGE+SELF	ENCOURAGE+SELF+SOCIAL	
Mailing Round	#1	2013-2014 absences 3+ greater than mode of classmates in school and grade			
	#2	N/A	All students in the ENCOURAGE group from Round 1	2+ absences as of data pull (otherwise ENCOURAGE mail)	2+ absences as of data pull AND missed more days than their modal classmate (If 0-1 absences, then ENCOURAGE mail. If 2+ absences, but have not missed more days than modal classmate, then ENCOURAGE+SELF)
	#3- #5	N/A	3+ absences as of data pull (otherwise, no mailing)	3+ absences as of data pull (otherwise, no mailing)	3+ absences as of data pull AND missed 2+ days more than their modal classmate (If 3+ absences, but have not missed 2+ more days than modal classmate, then ENCOURAGE+SELF)

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Primary Analysis

Impact on Days Absent

Our primary analysis assesses the impact of the treatments on days absent among students *in the main universe only*.⁷ We proceed along two parallel tracks. First, we report hypothesis tests of various sharp null hypotheses of no treatment effect. Second, we report raw and regression-adjusted group means and corresponding average treatment effects.

Hypothesis Testing

We use randomization inference to test various sharp null hypotheses of no impact. In particular, we define four potential outcomes for each unit:

- $Y_i(0)$: Unit i 's absences if assigned to the control group
- $Y_i(1)$: Unit i 's absences if assigned to the ENCOURAGE group
- $Y_i(2)$: Unit i 's absences if assigned to the ENCOURAGE + SELF group
- $Y_i(3)$: Unit i 's absences if assigned to the ENCOURAGE + SELF + NORMS group

where Y_i^{obs} corresponds to each unit's observed outcome. To use this potential outcomes notation, we assume that randomization was indeed valid and that the Stable Unit Treatment Value Assumption holds (SUTVA; Rubin, 1980).

Our primary null hypotheses of interest are the pairwise contrasts:

$$[H_{01}:] \quad Y_i(0) = Y_i(1) \text{ for all } i$$

$$[H_{12}:] \quad Y_i(1) = Y_i(2) \text{ for all } i$$

$$[H_{23}:] \quad Y_i(2) = Y_i(3) \text{ for all } i$$

and the global null hypothesis of no effect:

$$[H_{0123}:] \quad Y_i(0) = Y_i(1) = Y_i(2) = Y_i(3) \text{ for all } i$$

Covariate adjustment. Following Rosenbaum (2002) and Imbens and Rubin (2015), we regression-adjust the raw outcomes prior to testing. In particular, we covariate adjust by regressing Y_i^{obs} on the vector of control variables we defined above. Formally, we estimate the following regression:

$$Y_i^{obs} = \alpha + X'_{igs}\beta + \gamma_s + \delta_g + \varepsilon_{igs}$$

where the subscript is for student i in grade g in school s ; X is a vector of student-level controls (demographics, prior absences, etc.); and γ_s and δ_g are school and grade indicators, respectively.

⁷ The extreme and low absence universes are analyzed separately (see Additional Analyses).

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

The corresponding residual for each student is denoted r_i^{obs} . For consistency, we use the same residuals for all hypothesis tests and do not re-estimate this regression for comparisons that only use a subset of the data.

Pairwise null hypotheses. We test the three pairwise null hypotheses of interest against the one-sided alternatives:

$$[H_{01:}^a] \quad Y_i(0) \geq Y_i(1) \text{ for all } i$$

$$[H_{12:}^a] \quad Y_i(1) \geq Y_i(2) \text{ for all } i$$

$$[H_{23:}^a] \quad Y_i(2) \geq Y_i(3) \text{ for all } i$$

where at least one inequality is sharp. We obtain exact p-values using a Fisher Randomization Test with a one-sided Wilcoxon rank-sum test statistic.

Global null hypothesis. We test the global sharp null hypothesis:

$$[H_{0123:}] \quad Y_i(0) = Y_i(1) = Y_i(2) = Y_i(3) \text{ for all } i$$

against the one-sided alternative

$$[H_{0123:}^a] \quad Y_i(0) \geq Y_i(1) \geq Y_i(2) \geq Y_i(3) \text{ for all } i$$

where at least one inequality is sharp. We obtain an exact p-value using a Fisher Randomization Test with a test statistic of the minimum of the signed Wilcoxon test statistics from the three pairwise contrasts.

Multiple testing. We will use a multiple testing procedure to guarantee that the Family-Wise Error Rate for these tests is at most $\alpha = 0.05$.

Estimation

We will report both the raw and covariate-adjusted means for each treatment condition. Covariate adjustment consists of the regression of absences on the vector of treatment indicators, school and grade indicators, and the student-level covariates defined above. Specifically, we will fit the following model:

$$Y_i^{obs} = \alpha + \tau_1 T_{1igs} + \tau_2 T_{2igs} + \tau_3 T_{3igs} + X'_{igs} \beta + \gamma_s + \delta_g + \varepsilon_{igs}$$

where the subscript is for student i in grade g in school s ; τ_1 , τ_2 , and τ_3 are the coefficients of interest on the treatment indicators T_1 , T_2 , and T_3 , which correspond to ENCOURAGE, ENCOURAGE + SELF, ENCOURAGE + SELF + NORMS treatments, respectively. X is a vector of control variables we defined above; and γ_s and δ_g are school and grade fixed effects, respectively.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Sensitivity Checks

A practical concern is that the outcome distribution is likely to have a long right tail. We conduct two main sensitivity checks to assess this. First, we will repeat the above analyses defining the outcome at $\log(\text{Absence} + 1)$ rather than simply the raw number of days absent. Second, we will exclude all students with absences that are more than two standard deviations of their school-grade mean.

Another concern is that, due to discreteness, the observed probability of treatment assignment within each stratum differs slightly from the theoretical value of 0.25. The Fisher Randomization Test naturally addresses this complication, but the regression estimate does not. We therefore repeat the main analysis using Inverse Probability of Treatment Weights. Given the limited variability in treatment probabilities, we anticipate only modest changes from the un-weighted estimates.

“Sibling”⁸ Analysis

Some households had more than one student eligible for inclusion into our Main Universe. In those cases the focal student was randomly selected, which means we can analyze spillover effects on the non-focal sibling without bias.

For our main analysis, we limit our universe to two-student households where both students qualify for the Main Universe. Since each student had a 50% chance of being designated the focal student, the treatment effects among focal students and non-focal students are directly comparable.

First, we test whether there is any impact of randomization on the non-focal students. To do so, we repeat the primary hypothesis test procedure for non-focal students only. Second, we report the raw and regression-adjusted group means for the non-focal students, again repeating the procedure in the primary analysis. For comparability, we also report the raw and regression-adjusted group means for the *focal* students in two-student households, since these students are a subset of the main study universe.

Additional “Sibling” Analysis

While we have limited power to assess this, we hypothesize that there are differential impacts for two groups:

- focal student is in a higher or same grade as the non-focal student; and
- focal student is in a lower grade than the non-focal student.

We will assess this via an interaction term in the outcome regression.

⁸ Multiple students in the same household are not always necessarily siblings. We use the term sibling for simplicity.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Exploratory Subgroup Analysis for the Primary Outcome

We will assess impact variation across several subgroups. We view these subgroup analyses as exploratory and will assess each directly by including an interaction term in the outcome regression defined in the main analysis section.

Endogenous Stratification of Absences

We hypothesize that students with a higher propensity towards absenteeism are more strongly impacted by the treatment. To estimate absenteeism propensity, we use the endogenous stratification method of Abadie, Chingos, and West (2013), in which the control group is used as a training set to model absences in the 2014-2015 SY using covariates and absences in previous years. We then use this model to predict absence propensity in the entire sample universe. We will then look at treatment effects across absenteeism propensity terciles.

High School vs. All Other Grade Levels of School

We hypothesize that the interventions will differentially affect high schoolers as compared to K-8 students.

School-specific Communications

Since different schools have different protocols for communicating to guardians of highly truant students, we hypothesize that guardians of students with low levels of school-to-guardian communication will be more responsive to the treatment interventions, since the mailers are more likely to stand out from other communication. We test this hypothesis by drawing on data from the district's annual survey of guardians. We use this survey data to define schools as having relatively low to relatively high communication with guardians and interact this indicator with our treatment variable.

Analysis of Secondary Outcomes

We repeat the primary hypothesis test and estimation procedures for the following secondary outcomes:

- Number of tardies after postcard #1 was sent out through end of school year
- Grades of core courses averaged across the final three marking periods of the 2014-2015 SY
- State standardized reading and math test scores for all available grades

Missing Outcomes

Some students will leave the school district—either because they drop out of school or move away—between the beginning of the study and the end of the school year, creating missing outcomes. Since outcomes are collected from administrative data, we do not anticipate that the intervention will have a significant impact on missingness. As long as:

- there is no significant difference in missingness between experimental conditions, as determined by a simple F -test; and
- the overall proportion of missing outcomes is less than five percent of the total experimental sample,

then cases with missing outcomes will be excluded from the subsequent analysis (i.e., complete case analysis). If the proportion of missing outcomes is greater than five percent, then outcomes will be multiply imputed, assuming that outcomes are missing at random, conditional on the observed covariates, stratum, and treatment status. If missingness significantly differs across experimental conditions, then we will employ a principal stratification methodology, reporting effects for the Always Observed principal stratum.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Survey Analysis

We conducted a live phone survey of guardians in our experimental universes after Mailing 5 (between 6/20/2015-6/25/2015).⁹ The survey has two primary purposes:

- (1) *Internal Validity* - A set of questions address whether the guardians received, read, and understood the mail.
- (2) *Impact on Parental Beliefs* – How did the mail pieces impact parental beliefs about the importance of attendance and their role in ensuring their students get to school?

A secondary purpose of the survey is:

- (3) *Impact on Parental Behavior* – How did the mail treatments influence parental behavior?

Inclusions

We included all households in the main and extreme absence universes with:

- valid, non-duplicate phone numbers,
- students who were still enrolled in a school in our partner district,
- students who were not newly flagged with a Disability Indicator after random assignment.

Analysis Methods

Unless otherwise specified, all survey question analyses will be evaluated by fitting an OLS model with treatment indicators controlling for our standard battery of covariates.

Internal Validity

We included three survey questions that assess the internal validity of the treatment.

Recall: Since January, have you received letters or cards through the mail about [STUDENT FIRST NAME]’s attendance?

This question assesses whether the treatment interventions were successfully received. As such, we expect ENCOURAGE, ENCOURAGE + SELF and ENCOURAGE + SELF + NORMS to have significantly higher rates of recall as compared to CONTROL.

Relative Absences: This past school year, would you say that [STUDENT FIRST NAME], was absent from school more days than [HIS/HER] classmates in [HIS/HER] grade, about as many days as [HIS/HER] classmates in [HIS/HER] grade, or fewer days than [HIS/HER] classmates in [HIS/HER] grade?

With this question, we find whether the content of the treatment interventions was read and was successfully integrated into the guardians’ perception of their students’ absences. Only the NORMS component shows the comparison of the guardians’ student and the typical student, so we expect only ENCOURAGE + SELF + NORMS to exhibit a significant difference. Namely, we hypothesize that CONTROL, ENCOURAGE, ENCOURAGE + SELF are more likely to say “same” or “less than” as compared to ENCOURAGE + SELF + NORMS. Relatedly, we hypothesize that there will be no difference in responses across CONTROL, ENCOURAGE and ENCOURAGE + SELF.

⁹ For the full text of the survey, please see Appendix.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Since we have actual absence data about the focal student and the typical student in their school-grade, we can measure the accuracy of those social comparison beliefs. As before, we expect that there will be no difference across CONTROL, ENCOURAGE, and ENCOURAGE + SELF. We hypothesize that ENCOURAGE + SELF + NORMS will have greater levels of accuracy than CONTROL, ENCOURAGE and ENCOURAGE + SELF.

Absolute Absences: There were 180 days of school this year. On how many of those days do you think [STUDENT FIRST NAME] was absent from school, for unexcused or excused reasons?

This question similarly addresses whether guardians read and comprehended the information about their students' absences. However, this question looks at the content of the SELF component. Since prior research has shown that guardians tend to underestimate their students' absences on average, we first evaluate whether ENCOURAGE + SELF and ENCOURAGE + SELF + NORMS are associated with higher estimates as compared to CONTROL and ENCOURAGE. Since no information about the students' absences was included in the ENCOURAGE mailer, we expect no differences between ENCOURAGE and CONTROL.

As before, we include actual absence data to determine accuracy. In this case, accuracy is defined as the difference between actual absences and the number of absences guardians estimated or guessed. As before we expect no differences in accuracy across ENCOURAGE and CONTROL; we do expect that guardians in the ENCOURAGE + SELF and ENCOURAGE + SELF + NORMS conditions will be more accurate than guardians in the ENCOURAGE and CONTROL conditions.

Impact on Parental Beliefs

We included a battery of questions assessing parental beliefs about the importance of attendance and their role in getting their students to school.

Importance of Attendance: Thinking back to this past school year, which one behavior from the following list would have most helped [STUDENT FIRST NAME] do better in school?

(Caller reads options 1-6. List is randomized.)

- 1 Attend school each day
- 2 Do more academic work at home
- 3 Read everyday
- 4 Spend less time on phone, television, or electronics
- 5 Go to bed earlier
- 6 Eat breakfast

We hypothesize that the treatments will increase the likelihood that guardians will believe that attending school every day most helps their student do better in school. As such, we hypothesize that guardians in the ENCOURAGE, ENCOURAGE + SELF, and ENCOURAGE + SELF + NORMS conditions will choose answer choice 1 at a greater rate than guardians in CONTROL.

Parental Beliefs about their Role in Preventing the Students' Absences: (Multiple questions, please see questions Q12, Q16, Q19, Q20, Q14, Q15, and Q18 in the Appendix.)

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

We will use factor analysis to create indexes of related questions. Due to the potential of ceiling effects, we will not include questions where more than 85% of respondents in the CONTROL group select the same response. We expect that guardians in the ENCOURAGE, ENCOURAGE + SELF, and ENCOURAGE + SELF + NORMS conditions will have increased rates of belief that their role is important in preventing student absences as compared to guardians in the CONTROL group.

Parental Beliefs Divergent Validity: It is [STUDENT FIRST NAME]'s job to make sure [he/she] attends school every day.

Since none of the treatments emphasized that it is the student's job to make sure he/she attends school every day, we expect to see no differences between any of the treatment groups and CONTROL in terms of guardian responses to this question.

Parent/Student Relationship: [STUDENT FIRST NAME] and I have a warm and loving relationship.

We expect that the treatment mailings will not affect guardians' perception of their relationship with their student. As such, we hypothesize that there will be no differences between any of the treatment conditions and CONTROL in terms of responses to this question.

Impact on Parental Behavior

We will also evaluate whether there are other significant differences in survey responses across conditions, including whether guardians talked about absences, whether they are interested in receiving attendance information about their student in the future, whether they shared the mail pieces with other members in their household, and whether they communicated with the school about their students' attendance.

Additional Analyses

Additionally, we will evaluate the effectiveness of treatments in each of our secondary universes using the same methods described in the Primary Analysis section.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Appendix

Survey Logistics:

All surveys will be conducted in English.

Important Notes for the Phone Callers:

If questions arise during the phone survey, callers may reference the FAQ listed on pages 2 and 3 to respond to guardian questions. The callers may respond to questions with the answers as listed.

If the question is **not** listed on the FAQ's, callers should respond with:

"Thank you for asking. Sorry, I am unable to answer that question. I would suggest calling a member of the research team at the School District of XYZ at ABC. They will be able to answer your question."

If the parent/guardian has any questions about *the research study or phone survey*, respond with:

"Thank you for asking – for the questions you are raising, please call 123 to speak with a member of the research team."

If the parent/guardian has any questions about their *student's attendance data* respond with:

"Thank you for asking – for the questions you are raising, please call the School District of XYZ Attendance Office at ABC. They are open 8:30 am to 4:30 pm Monday through Friday. "

Survey Logic:

Question(s)	Notes
Q0	Parent/guardian verification
Q1-7	All households
Q8-11	If yes to Q7
Q12-20	Randomize order of questions
Q21-22	Gr 6-12 only
Q23	Gr K-5 only
Q24-29	All households Randomize list of options in Q27
Q30-32	Randomly ask 1 of these questions
Q33-38	All households

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

FAQ for the callers

Q: Who are you? Why are you calling?

A: My name is [CALLER NAME] and I am calling as part of the School of District of XYZ Attendance Project. The School District of XYZ is partnering with researchers at Harvard University for this project to study student attendance, parent beliefs, and academic performance. I am not a representative of the School District of XYZ. This is survey is for research purposes only.

Q: I already did this. Why are you calling me again?

A: OK. You may have participated in a similar call in January. This is a different call to collect some more information.

Q: Who is collecting/using the data from this phone survey?

A: The data will be used for the School of District of XYZ Attendance Project and is for research purposes only. The School District of XYZ is partnering with researchers at Harvard University for this project to study student attendance, parent beliefs, and academic performance.

Q: Are you a representative of the School District of XYZ or the student's school?

A: No, I am calling as part of the School District of XYZ Attendance Project. The School District of XYZ is partnering with researchers at Harvard University for this project to study student attendance, parent beliefs, and academic performance. The survey is conducted for research purposes only.

Q: Is my student in trouble?

A: No, this survey is for research purposes only. If you have questions or concerns about your student's attendance, please contact the district or your student's school directly.

Q: What is the purpose of this survey?

A: The purpose of this survey is to understand parent and guardian beliefs and behaviors related to student attendance.

Q: What does it mean for a parent/guardian to participate in this survey?

A: Participating parents and guardians answer the questions in this survey.

Q: Who is paying for this survey?

A: The Harvard research team is paying for this survey.

Q: If I take part in this survey, will the information you collect be kept confidential? How will our privacy be protected?

A: It is very important to us that any individually identifiable data is confidential. The data will be confidential and only used for this project. Only members of the research team and collaborators assisting with completion and analysis of the survey will have access to this data. The data will be publicly reported as group data – individual students or parents/guardians will not be identified. Only summary results will be shared publicly or with the School District of XYZ.

Q: What are the benefits and risks of this survey?

A: We see no likely risks or discomforts for you or your student. We cannot promise any benefits to you or your student from taking part in this survey.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Q: What is the purpose of this project?

A: The purpose of this project is to better understand student attendance, parent beliefs, and academic performance.

Q: Who is paying for this project?

A: The Harvard research team is paying for this project.

Q: Who is a part of this project? The parent/guardian or the student?

A: Both students and their respective parent and guardians are part of the project.

Q: What does it mean for my student to participate in this project?

A: Your student's participation in this project means student data regularly collected by the School District of XYZ will be shared with the research team, collaborators, and vendors (for example, a company that does mailings) helping with this project.

Q: What information about my student will be shared and collected?

A: The data includes attendance data, academic performance, demographics, and contact information.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

PHONE SURVEY SCRIPT

Q0. Hello. My name is [CALLER NAME] and I am calling as part of the School District of XYZ Attendance Project. May I please speak with the parent or guardian of [STUDENT FULL NAME]?

- 00 No, is not the parent/guardian
Thank you. This call is focused on [STUDENT FULL NAME]. Thank you for your time. Have a good [day/evening/night]. Goodbye. → *Terminate survey*
01 Yes, is the parent guardian → Go to Q1
90 Don't Know
92 Terminated

[student_attendance_number]

Q1. Thank you. The purpose of the project is to understand factors that influence student attendance. I am calling you today to ask you a few questions. You do not have to participate, and you may stop at any time.

First, I have a few questions about [STUDENT FIRST NAME]'s attendance in school. There were 180 days of school this year. On how many of those days do you think [STUDENT FIRST NAME] was absent from school, for unexcused or excused reasons?

[IF DON'T KNOW, ASK "IF YOU HAD TO GUESS, HOW MANY DAYS WAS YOUR STUDENT ABSENT?"]

- 1 Provided a number; Number is recorded (Q1a)
90 Don't Know → If gave a guess, number is recorded (Q1b)
91 Refused
92 Terminated at this question

[student_attendance_comparison]

Q2. Thank you. This past school year, would you say that [STUDENT FIRST NAME], was absent from school more days than [HIS/HER] classmates in [HIS/HER] grade, about as many days as [HIS/HER] classmates in [HIS/HER] grade, or fewer days than [HIS/HER] classmates in [HIS/HER] grade?

- 3 More days absent → Go to Q3
2 About as many days absent → Go to Q5
1 Fewer days absent → Go to Q4
90 Don't Know → Go to Q5
91 Refused → Go to Q5
92 Terminated at this question

[more_days]

Q3. Would you say that [HE/SHE] was absent MANY more days than [HIS/HER] classmates in [HIS/HER] grade, or JUST A COUPLE more days than [HIS/HER] classmates in [HIS/HER] grade?

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

- 2 Many more days → Go to Q5
- 1 Just a couple more → Go to Q5
- 90 Don't Know → Go to Q5
- 91 Refused → Go to Q5
- 92 Terminated at this question

[fewer_days]

Q4. Okay, thank you. Would you say that [HE/SHE] was absent MANY fewer days than [HIS/HER] classmates in [HIS/HER] grade, or JUST A COUPLE fewer days than [HIS/HER] classmates in [HIS/HER] grade?

- 2 Many fewer days
- 1 Just a couple fewer
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[communication_other_adults]

Q5. Over this past school year, have you spoken with the parent or guardian of any of [STUDENT FIRST NAME]'s classmates about how often their students are absent from school?

- 0 No
- 1 Yes
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[future_program_desire]

Q6. Please imagine that next school year you could receive mailings showing how many days of school [STUDENT FIRST NAME] had been absent, and also how many days [STUDENT FIRST NAME]'s typical classmate had been absent. Would you contact [STUDENT FIRST NAME]'s school to enroll so you could receive that information?

- 0 No
- 1 Yes
- 91 Refused
- 92 Terminated at this question

[received_letter]

Q7. Since January, have you received letters or cards through the mail about [STUDENT FIRST NAME]'s attendance?

- 0 No → Go to Q12
- 1 Yes → Go to Q8
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[letter_student]

Q8. Did you show these letters or cards to [STUDENT FIRST NAME]?

- 0 No
- 1 Yes
- 90 Don't Know

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

- 91 Refused
- 92 Terminated at this question

[letter_other_adults]

Q9. Okay, thank you. Did you show these letters or cards to other adults in your household?

- 0 No
- 1 Yes
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[throw_or_save]

Q10. Did you throw these letters or cards away as soon as you read them, or did you save them somewhere like on the counter or on the refrigerator?

- 1 Threw out the letters
- 2 Saved the letters
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[mail_or_text]

Q11. Would you have rather received this information through the mail or through text message?

- 1 Mail
- 2 Text message
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[parent_beliefs]

Thank you. Now I will read a series of statements. Please respond to the question "How strongly do you agree with each of the following statements?" Please use the following options: strongly agree, somewhat agree, somewhat disagree, strongly disagree. [ORDER of Q12-Q20 (inclusive) IS RANDOMIZED]

[role_construction1]

Q12. Making sure [STUDENT FIRST NAME] attends school every day is my responsibility.

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[divergent_validity2]

Q13. It is [STUDENT FIRST NAME]'s job to make sure [he/she] attends school every day.

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[Mechanisms_3]

Q14. Missing a few days of school each month has a huge effect on success in school.

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[mechanisms_4]

Q15. Absences are fine if I provide the school with an excuse or an explanation.

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[mechanisms_5]

Q16. Others think it is my job to make sure [STUDENT FIRST NAME] attends school every day.

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[Relationship2]ⁱ

Q17. [STUDENT FIRST NAME] and I have a warm and loving relationship.

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

[utility1]

Q18. Attending school each day is important for [STUDENT FIRST NAME] to succeed in school.

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[efficacy2]

Q19. I have influence over how many days of school [STUDENT FIRST NAME] misses.

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[invitations2]

Q20. The School District of XYZ and [STUDENT FIRST NAME]'s school want me to be involved in improving [his/her] attendance.

- 4 Strongly Agree
- 3 Agree
- 2 Disagree
- 1 Strongly Disagree
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

GRADES 6-12 ONLY

[6-12_schoolwork_time]

Q21. This past school year, how much time did [STUDENT FIRST NAME] spend doing school-related work at home on a typical night?

- 1 Less than 1 hour
- 2 1-2 hours
- 3 2-3 hours
- 4 3-4 hours
- 5 More than 4 hours
- 91 Refused
- 92 Terminated at this question

[6-12_best_students_schoolwork_time]

Q22. Okay, thank you. How much time do you think the students who get the very best grades in [STUDENT FIRST NAME]'s class spent doing school-related work at home on a typical night?

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

- 1 Less than 1 hour
- 2 1-2 hours
- 3 2-3 hours
- 4 3-4 hours
- 5 More than 4 hours
- 91 Refused
- 92 Terminated at this question

GRADES K-5 ONLY

[k-5_learning]

Q23. If you had to choose one, would you say that students in [GRADE] at [SCHOOL] this past year spent most of their time learning important skills or spent most of their time playing?

- 1 Learning important skills
- 2 Playing
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

ALL

[other_parent_contact_info_class]

Q24. Now we're interested in how much parents talk with each other. Do you know the name and phone number of a parent of another student in [GRADE] at [SCHOOL]?

- 0 No
- 1 Yes
- 91 Refused
- 92 Terminated at this question

[other_parent_contact_info_school]

Q25. Now, do you know the name and phone number of a parent of another student in ANY grade at [SCHOOL]?

- 0 No
- 1 Yes
- 91 Refused
- 92 Terminated at this question

[other_parent_school_transport_support]

Q26. Have you ever asked another parent of a student in ANY grade at [SCHOOL] for help taking your child to school?

- 0 No
- 1 Yes
- 91 Refused
- 92 Terminated at this question

[beneficial_student_actions]

Q27. Thinking back to this past school year, which one behavior from the following list would have most helped [STUDENT FIRST NAME] do better in school?

(Caller reads options 1-6. List is randomized.)

- 1 Attend school each day
- 2 Do more academic work at home
- 3 Read everyday
- 4 Spend less time on phone, television, or electronics
- 5 Go to bed earlier
- 6 Eat breakfast
- 7 Volunteered (Other)
- 91 Refused
- 92 Terminated at this question

[parent_teacher_conference]

Q28. Did you attend a parent teacher conference about [STUDENT FIRST NAME] this spring?

- 0 No
- 1 Yes
- 91 Refused
- 92 Terminated at this question

[absenceday1_reason]

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

Q29. Please think about the last time [STUDENT FIRST NAME] was absent from school. Which of the following best explains why [he/she] missed school that day?

(Caller reads options)

- 1 Asthma
- 2 Allergies
- 3 Illness of some other sort
- 4 Transportation Problems
- 5 Family emergency or event
- 6 Suspension
- 7 [He/she] woke up late
- 8 [He/she] did not want to go to school
- 9 You did not know that [he/she] missed school that day
- 10 Other reason
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

RANDOMLY ASSIGN TO Q30, Q31, OR Q32

[per_pupil_expenditure1]

Q30. How much money do you think the School District of XYZ spent per student this past academic year? Would you say the School District of XYZ spent:

(Caller reads options 1-7)

- 1 Less than \$3,000 for each student this year
- 2 \$3,000 for each student this year
- 3 \$6,000 for each student this year
- 4 \$9,000 for each student this year
- 5 \$12,000 for each student this year
- 6 \$15,000 for each student this year
- 7 \$18,000 for each student this year
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[per_pupil_expenditure2]

Q31. This past school year, the School District of XYZ spent around \$12,000 per student. Would you say that this is: more than you had thought, about what you had thought, or less than you had thought?

- 3 More than I thought
- 2 About what I thought
- 1 Less than I thought
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[per_pupil_expenditure3]

Q32. This past school year, the School District of XYZ spent around \$67 per student per school day. Would you say that this is: more than you had thought, about what you had thought, or less than you had thought?

- 3 More than I thought

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

- 2 About what I thought
- 1 Less than I thought
- 90 Don't Know
 - 91 Refused
 - 92 Terminated at this question

[communication_meeting]

Q33. Since January, have you had a meeting with the school or district about [STUDENT FIRST NAME]'s attendance?

- 0 No
- 1 Yes
- 90 Don't Know
 - 91 Refused
 - 92 Terminated at this question

[asthma]

Q34. Does [STUDENT FIRST NAME] suffer from asthma?

- 0 No
- 1 Yes
 - 90 Don't Know
- 91 Refused
 - 92 Terminated at this question

[allergies]

Q35. Does [STUDENT FIRST NAME] suffer from allergies?

- 0 No
- 1 Yes
- 90 Don't Know
 - 91 Refused
 - 92 Terminated at this question

[relationship]

Q36. Thank you. Just a few final questions. What is your relationship to [STUDENT FIRST NAME]?

(Caller reads options)

- 1 Parent
- 2 Grandparent
- 3 Aunt or uncle
- 4 Step parent
- 5 Sibling
- 6 Cousin
- 7 Foster parent
- 8 Other
 - 91 Refused
 - 92 Terminated at this question

[language]

Q37. What is the primary language spoken in your household?

- 1 English

Attendance Project in a Major Metropolitan School District
2014-2015 SY
Pre-Analysis Plan

- 2 Spanish
- 3 Portuguese
- 4 Chinese/Mandarin
- 5 Creole (Haitian)
- 6 Arabic
- 7 French
- 8 Other
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

[guardian_education]

Q38. Last question: what is the highest level of school you have completed?ⁱⁱ

- 1 None, or grade 1-8
- 2 High school incomplete (Grades 9-11)
- 3 High school graduate (Grade 12 or GED certificate)
- 4 Technical, trade, or vocational school AFTER high school
- 5 Some college, associate degree, no 4-year degree
- 6 College graduate (B.S., B.A., or other 4-year degree)
- 7 Some post-graduate or professional schooling, no degree (e.g. some graduate school)
- 8 Post-graduate or professional degree after college
(e.g., Master's Degree or Ph.D.; law or medical school)
- 90 Don't Know
- 91 Refused
- 92 Terminated at this question

Great, Thank you for your time! Enjoy your [day/evening/night]!

[PAUSE for 5 seconds]

ⁱⁱ PEW Research Center (2013). *Appendix D: Topline Questionnaire – October 2013 Higher Education and Gender Survey*. Retrieved from: http://www.pewsocialtrends.org/files/2014/02/higher-ed_topline.pdf.