

PPOL-560 Final Project

Lawand Yaseen & Phil Cork

12/16/2021

Part 1: Discuss Existing Research

1. Find a research paper on a topic of interest that uses OLS or probit/logit/LPM. Provide a complete citation to this work. In this paper, the authors investigate the effectiveness of additional messaging in aiding parents of low income students to verify as part of the process for enrolling in early child education in Orleans Parish, Louisiana.

Weixler L, Valant J, Bassok D, Doromal JB, Gerry A. Helping Parents Navigate the Early Childhood Education Enrollment Process: Experimental Evidence From New Orleans. *Educational Evaluation and Policy Analysis*. 2020;42(3):307-330. doi:10.3102/0162373720922237

The paper can be accessed through the following link: <https://journals.sagepub.com/doi/10.3102/0162373720922237>

1a. Discuss the key finding of the paper The authors of this paper implement a multivariate logit model to evaluate the effects differing types of communication had on whether students were successfully enrolled in public funded early childhood education programs. Working alongside district administrators, the authors created a randomized controlled trial in which some of the parents of a total 1,300 applicable students received the pre-existing, relatively modest means of communication from the district via email, some received these messages as well as weekly formal, officially worded text messages, and a third group received the pre-existing communications as well as weekly personalized text messages with a more conversational tone. The key finding of the paper is that when controlling for demographic variables, weekly text notifications increased verification rates by 7% regardless of whether they were personalized or formal in tone. The conversational tone did, however, increase engagement rates for some groups. While beyond the purview of this review, it is also worth noting the paper also conducted text analysis on communications with parents during the verification process, uncovering insights regarding the obstacles and questions parents had throughout. Finally, the paper notes that text messaging may be an attractive means of intervention with modest financial costs, with an estimated average cost of \$40 per additional verified student.

2. What is the analytical question? (If there are multiple analyses/questions addressed in the paper, focus on a single important question). In designing the trial, the authors pose three primary analytical questions. 1) Can text messages increase families' ECE verification and enrollment rates? 2) Are formal and personalized texts equally effective? And 3) Are the effects similar for pre-K and Head Start applicants? (Weixler et al. 311) While we will not focus on each of these questions in this analysis, it is worth considering the project holistically, as many of the specification and identification decisions made are driven by the multifaceted approach to this particular study. In particular, our analysis will focus on how and to what extent text messages, as well as their messaging, influence verification rates.

2a. Why is it important to answer this question? This is a crucial question to answer because research shows the importance of early childhood education in increasing both cognitive and emotional

development as well as improving level of achievements and life outcomes. Over the last two decades, public investment in early childhood education has increased, but the process remains an obstacle. As the authors note, “District officials identified verification as the key barrier to program access” for low income families who expressed a desire for early childhood education by initiating the enrollment process (Weixler et al. 308). Even when the ECE enrollment process has been streamlined through the OneApp program to allow for condensing multiple application processes into a single process, it still remains a deterrent to student enrollment. As many as 35% of applications that are submitted through OneApp did not verify and thus were not enrolled, despite an expressed interest and desire in doing so (Weixler et al. 308). More complex and involved than the previous steps in the process, the verification process includes presenting specific documents regarding personal identification, proof of residency, and verification of present income in person at specified locations and windows of time offered throughout the parish. In addition, this study is particularly relevant in that text messages have been shown by prior research to change behavior, thus providing a low-cost and easily integrated tool for increasing the verification and enrollment rate of students in early childhood education opportunities.

2b. Write out a statistical model and pay special attention to your subscripts, as these indicate your unit-of-analysis (e.g., individuals at a given point in time, or national data over time or perhaps individual data over time) To measure the effects of text correspondence on verification rates, the analysis was performed at a child level to mitigate the impact of a parent having more than one child in the experiment. The statistical model used to analyze the question discussed above was the following:

$$n_i = \alpha + \beta_1(\text{group}_2) + \beta_2(\text{group}_3) + \beta_3(HS) + \beta_4(\text{group}_2 * HS) + \beta_5(\text{group}_3 * HS) + \beta_J(\text{controls}) + e_i$$

Our outcome of interest, n_i is a dichotomous variable indicating whether or not an observation was verified. The two treatment groups who received messages on different platforms or received personalized text messages have their estimated effects shown by β_1 and β_2 respectively. Two interaction variables were also included to account for the impact of Head Start alongside the treatment groups. These are estimated in β_4 and β_5 . To use these, a dummy variable, “HS”, which demarcates whether or not a student applied for Head Start, is included in the model with the effect indicated by β_3 . Lastly, β_J represents the various estimated β 's determined for the forthcoming controls that were included to account for variation among the student population.

2bi. What is the dependent variable? The dependent variable of interest is whether or not the child was verified for ECE eligibility by the deadline outlined by the Louisiana Department of Education. The log odds of the outcome is denoted by n_i , which is reliant upon the treatment input along with relating variables that were also controlled for. It is important to note that log odds differentiates from probability which is the method of interpretation we are more familiar with. Odds represent the proportion of successful observations to the failed observations as opposed to the definition of probability as the ratio of success to total observations. In this study, the log of the observed odds was taken.

2bii. What is the key independent variable (or, in some cases, variables)? The primary independent variable of interest was a set of three treatments pertaining to the communication format that the parents of the children received. Three groups were created and randomly assigned, with the first group, referred to as Group 1, receiving the standard weekly reminders in email and up to 5 weekend texts sharing verification reminders. Group 2 additionally received weekly text messages which maintained formal communication style. Also, Group 3 also received weekly text messages but had personalized content within each message. Within the statistical model, Group 1 was the reference variable while Groups 2 and 3 were included as dummy variables indicating treatment. In addition, children were also able to complete an application for Head Start. Head Start is designed to only be available to families who report a household income either at or below the national poverty line. As there are eligibility requirements that are more stringent than the general ECE process, it is assumed that the process is more difficult and therefore more of a deterrent to completing verification. To account for this, this was interacted with being in a respective treatment group. Lastly, a number of controlling demographic variables were included in the model in an

attempt to draw further conclusions about the effects of the treatment on subsets of the population. As discussed later on, these included income, whether or not the primary language spoken was not English, child's age and gender.

2biii. What are the null and alternative hypotheses? While there is no explicit outline of the null or alternative hypothesis, it can be derived that the null hypothesis is that the novel forms of communication do not make a meaningful difference to odds of verification. In this scenario, the alternative hypothesis is that either text messages or the personalization of these text messages will have some form of impact on the odds of verification. In statistical terminology, the null hypothesis is that $\beta_1 = \beta_2 = \beta_4 = \beta_5 = 0$ while the alternative hypothesis is that $\beta_1, \beta_2, \beta_4, \beta_5 \neq 0$.

2c. Does the model use any specific functional form (e.g., probit, log)? If yes, explain why. The model used by the authors employs a logit model. As the outcome of interest, verification, is reported as a dichotomous variable, use of any other model would not be logical. As we have discussed in class, there is little to no difference in using either probit or logit. The authors' use of the logit rather a probit model implies that they are assuming that the error term is distributed logistically rather than being distributed normally.

3. Endogeneity Because the test is a random chosen trial, any correlated variables left in the error term are inherently random in their relationship to the independent variables, thus preventing endogeneity from affecting the causal inferences drawn by the model's results. Thus, even variables one could think of as 'left in the error term' that might be correlated with an independent variable of interest and influence the dependent variable would not be a reason to be concerned about bias in the model affecting the causal inference in the outcomes.

3a. Which sources of endogeneity are the authors able to control for with additional control variables? The authors are incredibly interested in removing as much endogeneity as possible from the experiment. As seen in the "controls" which are included in the statistical model, the verification data included the following informative demographic categories: income, whether or not the primary language spoken was not English, child's age and gender. Of these, the two most likely variables to cause endogeneity would've been income and the primary language. As we'll discuss further, the premise of ECE is to provide childcare for families who are underneath a certain threshold of income. However, since the experiment is an RCT, the treatment will not be correlated with the reported income and will not yield endogeneity. Similarly, there can be conclusions drawn about the potential impact of not speaking English. As the verification process is in English, the difficulty to complete it would be related to a parent's ability to speak English. Once again, as the treatment is done using an RCT, there will not be endogeneity.

3b. Which sources of endogeneity are the authors unable to control for? Explain the implications for your analysis (only work through one such (important) example in detail). Use the conditions for omitted variable bias and our discussion of the implications of omitted variable bias to discuss whether the particular omitted factor will bias results upward or downward. If the paper is a RCT, explain a source of bias that the RCT is designed to avoid. As previously mentioned, there is not much room for endogeneity since the experiment is conducted using an RCT. However, there are some variables that the authors chose not to include in the statistical model as part of their controls that are worth mentioning. First, the work schedules of the parent's involved in the study were not taken into account in the treatment groups. While there is no specification about when the messages were sent, we will assume that it was automated to be sent at the same time. There could be some correlation between the parent's availability when receiving their text message to the verification of their student's eligibility. While this will be more abstract to quantify, it could lead to some endogeneity.

On the other hand, the RCT in its current form is best designed to avoid any endogeneity from household income. In an observational study, it would've been likely that income would be correlated with probability

of verification. However, since the treatment is applied randomly, the effects of the treatment groups will not be correlated with the individual's income.

3c. Could measurement error cause endogeneity? Explain why or why not. When considering measurement error, it is important to distinguish between two possible instances where error in the recorded values for variables could cause an issue. Were the measurement of our dependent variable, student verification in this instance, to be erroneous, this wouldn't cause bias or endogeneity as the inaccurate variation in the variable would simply be captured by the error term in the model. If one of our independent variables were measured incorrectly or inconsistently, however, this could be cause for concern, as it would cause attenuation bias. Attenuation bias causes the $\hat{\beta}_i$ value for X_i to be closer to 0 than it otherwise would be, essentially misrepresenting and under-reporting the magnitude of the relationship.

In this experiment, it is worth noting the income variable is self-reported by the parents applying on behalf of their students. There are three notable opportunities for measurement error with this feature in particular. Most prominently, there is a broad incentive to over-report income due to social pressure and norms. In applying for early childhood education programs, there may also be an incentive to under-report income so as to qualify for Head Start, which requires the family to be living 100% below the poverty line. For other programs. Finally, these two considerations may be mitigated by the verification process itself and potential penalization such as time lost in enrollment process or other consequences should the misreported income be challenged by an administrator of the application process. Finally, this same logic applies to much of the other demographic features included in the model — if there were to be measurement error, it would be erroneously verified by those managing the OneApp system and thus would be consistent within the model itself.

4. Post-treatment: Post-treatment: are any of the included variables possibly “post-treatment” variables? Because the included variables are almost exclusively demographic in nature, they inherently pre-exist the treatment variable of which messaging a particular parent received. Which randomly assigned group an individual observation was assigned could have no indirect influence on the student's gender, age, their parent's income level, or program type, at least not so beyond the bounds of those who self-selected into applying for ECE in the first place. Thus without a causal link between the randomly assigned control groups and one of the other independent variables included in the model specification, it is unlikely any of the included variables would be considered “post-treatment” when evaluating the model for potential biases.

5. What factors affect the precision of the results? The primary means of evaluating precision of the results is to consider the variance of the estimated coefficients. This variance is in turn affected by the variance of the regression itself, the sample size, the variance of the independent variables, and the degree to which the independent variables are correlated with one another. With regard to multicollinearity, there is likely to be an observable degree of correlation between income and whether a child applied for a Head Start program given the prerequisites for Head Start being based on income. With a sufficiently large sample size as detailed above, and a presumed naturally diverse variance of the independent variables, the precision of the model is not put in doubt based on the variance of the estimated coefficients.

5a. Discuss, as appropriate autocorrelation, heteroscedasticity and multicollinearity and what the authors do (or should do) about them. Focussing only on whether a student is verified or not, autocorrelation is not an issue in this hypothetical experiment, as there is no temporal element in which certain observations would provide information about the error term of other observations. Heteroscedasticity is also unlikely to be a prevalent issue affecting the standard errors of the model. Multicollinearity is perhaps more likely to be of concern in this iteration. With the Head Start, and by extension the interaction variables composed of the dichotomous Head Start variable, having an inherent relationship with income, will lead to inevitable multicollinearity. This effect will not add bias to the model, but will lead to higher standard errors in the model's results.

5b. Is the test statistically powerful? What are the implications of your answer to this question for how you will interpret your results? To be statistically powerful, the data and accompanying model must be equipped to limit the possibility of type II error. As the authors use a variety of levels of statistical significance in their results as shown on page 318, the most straightforward to to minimize type II error is to decrease ($se(\beta_1)$) as much as possible. As standard error and the number of observations are inversely related, a higher N will yield a relatively lower standard error. In this experiment, the number of observations for each treatment group was roughly 1400. Given the size of the treatment groups, we can assume in our analysis that the authors' likelihood of making a type II error is minimized and any conclusions drawn about potential causal relationships are statistically sound.

6. Have the authors considered heterogeneous treatment effects? (That is, have they considered whether the effect of the key independent variable(s) differs across sub-groups?) Describe how they do or could address this possibility. To adequately draw conclusions while taking into account heterogeneous treatment effects, the authors' included two interaction variables, both of which investigate the impact of needing to verify for Head Start along with the treatment groups. The reasoning they outlined for inclusion of the interaction variables was "the verification process in New Orleans is more complicated for Head Start applicants and because this could be a particularly disadvantaged population" (Weixler 316) As Head Start applicants are eligible due to being below an income threshold, they are required to provide more information to verify their eligibility. Since it is a more thorough process, the expectation was that the effect of different communication methods could have a compounding impact on the subset of the population eligible for Head Start. For the population not eligible for Head Start, the inclusion of these variables will not have an impact on their log odds of eligibility as the interaction variable will be set to 0, nullifying the estimated coefficients β_4 and β_5 .

7. Are the results generalizable? Throughout the paper, the authors go to great lengths to emphasize the specific circumstances under which these results were observed. These findings represent the outcomes for a specific time in a specific place. While none of the variables considered are particularly unique to Orleans Parish, it is worth considering for reasons not considered here, results may vary geographically and temporally. A means of further testing the reproducibility and general interpretations of this study would be to incorporate these communication techniques in other cities utilizing the OneApp program like Chicago, Washington D.C., and others.

Part 2: Discuss a new research design

Develop a hypothesis to test on a question related to the above article. This hypothesis must be different than the hypothesis tested in the paper discussed in part 1.

1. What is the analytical question? While the experiment shows that additional messages and increased use of novel mediums such as text messages can have a beneficial result on verification rates of low-income students for early childhood education, the treatment does not address the main obstacle of the verification process itself - the arduous process of bringing identification and proof of income documentation to specific locations at designated times. With other municipalities offering the ability to verify through digital communication, typically with photos of documents and via phone interview, there's a precedent for more intuitive and less demanding verification processes. With this in mind, a follow-up experiment would be to see if individuals randomly selected to participate in a digital-only verification process were more likely to verify for ECE.

1a. Why is it important to answer this question? This question is crucial to explore for potentially providing easier access for low income students, who would arguably benefit from these additional opportunities the most, to verify for ECE programs. It could also lead to the process being more equitable as well, since an option not requiring in-person or time-specific steps would allow parents whose work schedule

doesn't allow the necessary flexibility to participate in an alternative means of completing this crucial stage of the enrollment process. With a program designed to provide services to low income families who are also likely to have the resources or experience to take advantage of them, iterative improvements of the process are of the utmost importance.

While this experiment would have additional execution challenges because it can not be as easily integrated into the existing process without additional administrative overhead, it is also potentially a more influential experiment on improving the entire process in a more holistic and equitable way that would pay long-term dividends for both the community and Parish administrators. In addition, the effects of early childhood education on higher education outcomes have been heavily documented and if this experiment yields the anticipated results, it would point to a method for the city of New Orleans to improve education outcomes with a simple administrative process.

Upon reflection of the initial results, the authors wrapped the obstacles of verification into three groups: awareness, understanding and capacity. The verification process presents difficulties of its own that exemplify the lengths to which a parent must understand in order to confirm their children's eligibility. While the initial experiment importantly addresses the awareness and to a lesser extent, the understanding aspects of these challenges, this hypothetical experiment would target the challenge of capacity for interested families.

1b. Write out a statistical model and pay special attention to your subscripts, as these indicate your unit-of-analysis (e.g., individuals at a given point in time, or national data over time or perhaps individual data over time). To measure the effects of a virtual-only verification process, we will employ an RCT. The model to perform this analysis is shown here:

$$n_i = \beta_0 + \beta_1(treatment) + \beta_2(HS) + \beta_3(HS*treatment) + \beta_4(schedule) + \beta_5(employment) + \beta_5(controls) + e_i$$

The outcome, n_i , is the log odds of a child being verified given the forthcoming variables. Our treatment's effect will be conveyed by β_1 , while the combination of β_2 and β_3 represent the effect of Head Start coupled with an interaction variable based off the treatment. We will be adding in two more controlling variables, accounting for the parent's work capacity and their employment status, shown by β_4 and β_5 respectively. We will include the same SES controls that the previous author included in their model that investigated verification.

1bi. What is the dependent variable? The dependent variable remains the verification of students who are eligible for ECE programs their parents have applied for previously through the OneApp program. This dependent variable is a dichotomous variable, capturing only if the child is verified prior to the deadline. It will follow the traditional dichotomous variable structure, where children who are verified will be assigned a 1 while those who are not will be assigned a 0.

1bii. What is the key independent variable (or, in some cases, variables)? This experiment would largely follow the model specification of the previous trial. The independent variables would include demographic information, but would not feature the treatment groups nor the Head Start interaction variable for them. By an extension of the author's analysis, Head Start would share an interaction variable with the treatment variable to control how applying for Head Start, a more arduous process, might affect the verification rates of those randomly assigned to the digital verification process. With this new approach considered, this experiment would also take into account two new variables to measure capacity and past experience with the system.

For measuring capacity, information would be gathered regarding the parent's work schedule and employment status. This input would help identify whether the changes were helping those who would have had more issues using the current system. Past experience with the system would track if a student had an older sibling who had gone through the enrollment process, which would highlight a parent's pre-existing knowledge of the system, potentially making it more likely they would complete the process regardless of digital or in-person

processes. This variable also potentially helps identify where the awareness and understanding challenges are overcome, if capacity is an issue in more isolation.

1biii. What are the null and alternative hypotheses? The null hypothesis for this experiment is that providing a digital-only process for verification does not have a meaningful influence on whether students verify prior to the deadline.

The alternative hypothesis is that there is a non-zero relationship between providing a digital-only process for verification that is statistically and substantively significant.

1c. Does the model use any specific functional form (e.g., probit, log)? If yes, explain why. As with the prior experiment, the model will implement a logit form due to the dichotomous dependent variable. While logit models provide less direct interpretation than a linear probability model, the s-shaped curve that results from the logit model can more accurately capture the bounded probabilities between 0 and 1 as well as the non-linear relationship between the independent and dependent variables. In addition, a linear probability model has a potentially flawed structure for this experiment - extreme predicted values below 0 or above 1 would show a lower coefficient of effect, even though those predictions point to a stronger relationship. For these reasons, we will employ the use of a logit model while assuming the error within the model is logistically distributed.

2. Endogeneity By randomly assigning individuals who have already applied to ECE programs to one of two verification processes, the experiment is able to sever the ties between unmeasured and correlated variables left in the error term and the variables included in the specification.

2a. Which sources of endogeneity would you expect to be able to control for with control variables? By incorporating broad demographic features, as well as indicators of a parent's availability to engage with the process via their work schedule or level of employment and their past experiences with the system, the model is able to control for the inherent variance of individuals in each treatment group as well as their ability to tackle possible endogeneity that would rise from not adequately addressing the three challenges of the verification process, awareness, understanding, and capacity through model specification. In addition, the use of an RCT carries the brunt of work of avoiding endogeneity as the randomized treatment will not be correlated with any other variable that may potentially be correlated with the outcome of interest.

2b. Which sources of endogeneity would you expect to be unable to control for? Explain the implications for your analysis. Use the conditions for omitted variable bias and our discussion of the implications of omitted variable bias to discuss whether the particular omitted factor will bias results upward or downward. If you propose a RCT, explain a source of bias that the RCT will avoid. (That is, for a RCT paper, explain a source of endogeneity that would occur if a RCT were not used.) The design of this RCT directly attempts to avoid the possible endogeneity of a parent's degree of financial or schedule flexibility and their pre-existing understanding of the system's process. Both of these indicators, were the treatment was not randomly assigned, would likely be correlated with one or more of the included demographic variables and would have a reasonable influence on the dependent variable of whether a student is verified or not. For example, by taking into account work capacity, we are hoping to eliminate the potential endogeneity that may arise from parents who work atypical schedules and may be unavailable or not respond to messages during expected hours. What we could assign as a potential relationship between income or any particular demographic, could in reality be more closely aligned with a parent's availability.

2c. Could measurement error cause endogeneity? Explain why or why not. This experiment has largely the same degree of potential for measurement error as the previous experiment. Likewise, there

are social and cultural pressures to potentially misrepresent information regarding income, or even to under-report to qualify for the ECE or Head Start programs, but because all of the reported information is to be verified by the system, there's a potentially greater incentive to not misrepresent any of the information. Had there been self-reported income, since it serves as an independent variable in our model, then there would potentially be measurement error which would've led to attenuation bias.

3. Post-treatment: are any of the included variables possibly “post-treatment” variables?

Because all of the control variables predate the randomly assigned treatment independent variable of interest, there are not included post-treatment variables in this particular experiment.

4. What factors affect the precision of the results? Given the experiment's similar specification to the previous iteration, the precision of the results is largely affected by the same inputs that impact the variance of the estimated coefficient of interest. Namely, a sufficiently large sample size, a suitable degree of variance in the independent variables, the variance of the model itself, and the included multicollinearity are the primary factors to consider when evaluating precision. While there may be some mild considerations for multicollinearity in this model specification, the sufficiently large sample size and inherent variance of demographic information as independent variables from randomly selected individuals is presumed to be more than enough to overcome these potential increases in the variance of the estimated coefficient.

4a. Discuss, as appropriate autocorrelation, heteroscedasticity and multicollinearity and what you would do about them.

Without a temporal component to the model, it is improbable to consider autocorrelation as a threat to the validity of the model's reported standard errors. Heteroscedasticity is also likely to not influence the model, as there's no prevalent reason to assume the errors would not be relatively uniformly distributed. As noted above, multicollinearity is a potential issue with multiple variables loosely capturing similar information regarding income. This reality will not cause bias in the model, but rather inflate the variance a bit. With the sufficient sample size and likely sufficient variance in the other independent variables, this effect is worth noting, but not likely to invalidate or deteriorate the quality of the results.

4b. Is the test statistically powerful? What are the implications of your answer to this question for how you will interpret your results?

The test would likely be just as, if not more so, powerful than the previous experiment. With only two treatment groups instead of three, each subset of the sample would be better represented in terms of the size of N. A higher power model would allow for more fine-tuned interpretation of results, particularly correctly rejecting or failing to reject the null hypothesis based on estimated coefficients that may be closer to zero, but still statistically significant, which is easier to determine in sufficiently high-power models.

5. Do you expect the treatment effects to be heterogeneous? (That is, will the effect of the key independent variable differ across sub-groups?) Describe how to address this possibility.

While our main treatment will be applied using an RCT, we do not anticipate the effect to differ among the traditional demographic subgroups. However, since there will be an interaction variable between eligibility for Head Start and the anticipated treatment, we expect that there will be a heterogeneous effect for those who are able to verify through a virtual verification process. That is by design and should inform us about the relationship between income eligibility and barriers that exist to verifying ECE eligibility.

6. Are the results generalizable?

This experiment would likely require preemptive qualifiers that the results, whether statistically significant or not, were drawn from a RCT working with a specific demographic, in a certain location, in a certain window of time. When considering a treatment that is larger in scope by changing the system demonstrably rather than adding additional touch-points of communication, this hesitation to assume results would directly correlate to other iterations of the experiment is warranted. With that said, there are no particular factors that stand out as setting low income students from Orleans Parish so

distinctly apart from comparable communities that would rule out the results being fairly applicable either, so long as precautions were made regarding any further assumptions based on the results.