List-based organizing and polling place vote tripling in the Ohio primary

Ken Stanley, kendallstanley.com
Oliver A. McClellan, Columbia University
Gabriel Zucker, Director of Research, Vote Rev
June 22, 2020

Executive Summary

Prior to Ohio's postponed March 17th primary, Ken Stanley conducted a randomized evaluation of list-based relational organizing and polling place vote tripling (PPVT). A list of 1,100 Oberlin College students who were registered to vote was randomly divided into a treatment group of 900 and a control group of 200. The list of 900 was shared with hundreds of Oberlin College students over the course of several days, primarily using a PPVT model, in which students were given the list and encouraged to remind three friends to vote as they left the early voting location.

In just two days of PPVT organizing, over 200 students agreed to send voting reminders, a very large number compared to other relational organizing programs in recent years. Notably, this group of students was not a disproportionately "activist" crowd, with many irregular voters agreeing to send reminders.

We estimate that turnout in the treatment group is 7.8 percentage points higher than in the control, though alternate specifications produce estimates ranging from 7 to 10pp.

The implied impact per recruited vote tripler is .1-.3 net votes. Despite the small sample size, most specifications are significant at $p<.05$. There is suggestive evidence that most of the treatment effect is driven by increased turnout among female voters.

There is an unexpected temporal pattern in these results, with a large portion of the effect occurring after random assignment but before the treatment had been fully implemented. It seems likely that some of the overall treatment effect was simply due to good luck. However, even if the first two days are dropped entirely from the experiment, estimated effect sizes are still around 4.5-6.5pp and, despite the small remaining sample size, estimates nearly reach traditional levels of statistical significance.

The entire experiment was designed and conducted within a period of three days, at a cost of $35 in materials and 21 person-hours of organizing. Even if organizers were paid $20/hr, the entire program would still cost a miniscule $6.50/net vote. For various reasons, the external validity of this study may be low, with similar implementations in other contexts likely showing smaller effect sizes. But, even if the true effect were as low as 25% of that estimated here, this effort would still represent one of the most cost-effective GOTV interventions analyzed in recent years, at $26/net vote.

1. Background

Recent years have seen increasing interest in relational organizing (RO) (see e.g., Han 2016, Schein et al. 2021, Green and McClellan 2020). This interest seems to be driven by RO programs' effectiveness. For instance, Green and McClellan (2020) recently found that Turnout Nation's 2019 RO campaign produced the largest intent-to-treat effect documented by any experimental GOTV study in the public domain in the past two decades.

The challenge for RO programs has always been scalability: While friend-to-friend contacts are powerful, producing large numbers of them is difficult. Most programs to date require supporters to download apps, receive extensive coaching from campaigns, or both. Mobilizing large numbers of people to opt into these intensive programs has proven challenging.

Two recent innovations have been gaining traction to overcome the RO scalability problem. The first is Polling Place Vote Tripling (PPVT). Vote Tripling in general is a tactic that theorizes plenty of voters *are* willing to do some relational organizing, as long as the ask is bite-sized. Specifically, if you ask voters to remind just three friends to vote, many will agree to do so, and many will follow through. Those who agree to remind three friends are referred to as "triplers." The conversion rate has been shown to be especially high when the ask is made of voters exiting the polling place after voting (thus Polling Place Vote Tripling), when over half of voters can reliably be convinced to text a few friends.

The second is browse-based RO. Most RO programs to date derive their list of a voter's friends from the voter herself — either by syncing her phone contacts, or having her think of a few contacts from memory. In *browse* RO programs, voters are shown a list of nearby targets, and asked which of them they know. The browse paradigm is powerful in getting voters invested; people are less reluctant to identify their friends when presented with a list and indeed are generally proud to identify the names they recognize. While they may know only a few names on the list, voters are generally excited to browse it, and enthusiastic to follow through when they recognize a name.

The experiment described in this paper is an amalgam of these two ideas: a unique implementation of PPVT, in which triplers were encouraged to select their targets off a list of voters who had not yet voted. Unlike traditional vote tripling tactics, in which voters select three arbitrary friends who cannot generally be mapped to the voter file, this browse-tripling hybrid has an impact on a known group of people, and thus can be rigorously analyzed. In this experiment, a randomly selected subset of voters was placed on the browse list, and vote triplers were recruited outside the polling place to contact on-list voters. (The target list was also circulated via other modes, which ultimately represented a relatively small portion of the overall program; see Section 2.) By comparing the turnout of on- and off-list voters, we can estimate the impact of the friend-to-friend reminders.

The balance of this paper is organized as follows. Section 2 outlines the experimental design and implementation. Section 3 describes the sample. Section 4 analyzes which voters choose to be triplers. Section 5 presents the principal turnout results. Section 6 highlights limitations to the study's external validity and concludes.

2. Experimental Design and Implementation

The experiment was conducted among Oberlin College students during early voting leading up to Ohio's scheduled presidential primary on March 17. Due to quirks of the quickly-moving coronavirus pandemic, early voting suddenly became the only way to vote in the primary available to Oberlin students living on campus. On Thursday, March 12, Oberlin students — the vast majority of whom live on campus — were informed they would have to vacate their dormitories[1] by noon on Monday, March 16, the day before Ohio's scheduled primary. As most Oberlin students are not local, this meant that most students registered on campus would have to vote during early voting on March 13, 14, 15, or 16 if they planned to cast their ballot at all. Unlike the election day polling location, however, the early voting location for Oberlin students is inconvenient, situated a 20-minute drive from campus, where many students do not have cars. Thus it is unlikely campus-registered students would have voted before March 13, and unlikely they would be able to vote after March 16. (Ultimately, in-person voting on March 17 was canceled at the last minute, and the deadline for absentee ballots to arrive was extended another six weeks: but this was not known at the time.) As such, for most on-campus voters, these four days became, in a sense, election day.

On March 12, using public voter file data, Stanley created a list of 1,100 Oberlin College students registered to vote in Oberlin and likely to be current voters at that location, who had not[2] yet voted in the primary. 900 of these were randomly assigned to treatment, and placed on the tripling target list. 200 were assigned to control and not included on any lists.

The list was disseminated to students in two distinct ways. First, on Friday, March 13, the list was circulated in a Google Spreadsheet through a variety of student channels, especially via a network of ride-shares that had sprung up to bring students to the early voting location.[3] Second, on Saturday and Sunday March 14 and 15, the list was disseminated via PPVT. Stanley stood outside the early voting location 2 hours Friday, 8

[1] https://www.oberlin.edu/campus-resources/bulletins/departure-campus-update-boxes-and-dining

[2] Currently registered students were defined as any students who had either voted in November 2019, or whose voter file entry had been updated since the spring of 2019. Oberlin College students are easily identifiable in the voter file as essentially all Oberlin College students use an Oberlin College Mail Room address as their mailing address, and so all voters with OCMR in their file can be categorized as Oberlin students. Ultimately, 20 students in the sample universe had already voted by the time the intervention began to be implemented.

[3] Stanley printed and gave the list to one of the drivers on Friday. This driver, whom Stanley has known for years, said that he showed it to the 24 students that he drove to the polls Friday afternoon. Stanley also shared with other drivers electronically. He has no evidence that they used the list.

hours Saturday and 5 hours on Sunday, with the treatment list of student voters[4]. As students exited the polling place, Stanley asked them to commit to finding a few friends on the list and reminding them to vote. Anecdotally, the larger portion of the effect appeared to be generated using PPVT, rather than the earlier methods of list dissemination, although this hypothesis cannot be directly tested.

More details on the PPVT model are available from Vote Rev or from Ken Stanley.

3. Sample and voter file data

The randomization was based purely on public voter file data, although we also purchased Targetsmart data on the sample for endline analysis. Of the 1,100 voters in the subject pool, Targetsmart matches were found for 988 of them. Table 1 shows descriptive statistics and a balance check based on the 988 matched records. Row 14 is based on analysis of registration[5] addresses rather than Targetsmart data, and so has n=1,100. Note that detailed voter file data were not available for the initial randomization, so there was no stratification or ex ante balance check performed.

While the randomization produced relatively balanced groups, some imbalance remains: treatment students have voted more regularly in past elections (with the notable exception of 2019), first registered longer ago (despite being only very slightly older), and are predicted to be somewhat more Democrat-leaning. The stronger voting history shows up in Targetsmart's[6] turnout propensity scores, though perhaps somewhat more weakly than one would expect. A priori, we would expect the treatment voters to turn out at rates a few percentage points higher than control voters. As such, it is prudent to run main results with controls as a robustness check.

[4] On Saturday, the list was in a Google Sheet, which was shared to students' phones as they left the polls; on Sunday, the list (growing shorter each day as students voted) was printed on two sides of a single sheet of paper. Students were encouraged to mark the names they knew, took the sheet with them, and were encouraged to text those friends on the ride back to Oberlin College.
[5] Voters living at an address with more than 14 Oberlin students in the full voter database are classified as on-campus.
[6] This data was extracted from Targetsmart on March 31, 2020, which means that in theory Targetsmart's turnout and partisanship scores *could* be endogenous. Targetsmart's documentation page, however, as of late April 2020, noted that no predictive models had been refreshed or rebuilt since December 2019. Still, for robustness, regressions with controls are also run solely with demographics and vote history variables that could not have been influenced by March 2020 voting behavior.

Table 1: Descriptive statistics and balance check

		Control mean	Treatment mean	t
1	Voted 2016 general[7]	.217	.266	1.37
2	Voted 2017 general	.05	.073	1.10
3	Voted 2018 general	.467	.527	1.47
4	Voted 2019 general	.339	.292	-1.24
5	Ever voted	.694	.713	.49
6	Age	20.8	20.9	.63
7	Percent female[8]	.556	.517	-.93
8	Minority	.067	.100	1.40
9	Presidential general turnout score (0-100)	56.5	57.0	.54
10	Off-year general turnout score (0-100)	14.4	15.4	1.25
11	Presidential primary turnout score (0-100)	13.1	13.5	1.16
12	Partisanship score (0-100)	96.4	96.7	1.45
13	Earliest registration date	7/12/18	5/14/18	-1.53
14	Lives off campus	.090	.109	.79

4. Tripler sign-up

Before turning to the results of the randomized study, it is worth observationally noting the extremely high vote tripling recruitment rate. In approximately 15 hours of work outside the polling place, Stanley recruited an estimated 250 students — about three quarters of the[9] students who voted during that period — to remind their friends to vote.

[7] This is coded as zero for anyone who did not vote, regardless of whether or not they were eligible or registered to vote at the time of the election.
[8] Some genders are unknown. Male and unknown are both coded as zero.
[9] Our analysis of voter file data suggests about 340 Oberlin students voted in person during the days Stanley implemented PPVT.

Most agreed to remind[10] several friends, with five the number most frequently reported. To compare the tripler recruitment rate to other relational programs, see this article.

Unlike when recruiting vote tripling pledges by canvass, phonebank, or textbank, the PPVT design does not generally allow us to determine who agrees to vote triple. However, a quirk of implementation does allow us partially to do so: For part of the experiment, some voters who[11] pledged to triple received the list via a Google Spreadsheet, and their emails are stored as collaborators on that document. By matching the names on Google Accounts to the voter file list, we can derive a partial list of pledged vote triplers.

One hundred thirty-eight (138) accounts were shared on the spreadsheet, of which 96 were matched to the Oberlin[12] College students on the original list of 1,100, i.e. the sample frame. All but 1 of these 96 is recorded as having voted during the experimental period, suggesting that the matches are generally valid. These 96 individuals are only a subset of the approximately 200-250 recruited triplers, but they are sufficient to draw some conclusions about who agrees to vote triple.

Table 2 shows the result of several logistic regressions predicting tripler opt-in based on a variety of characteristics from the voter file. Only students who actually voted during early voting prior to March 16 are included in the sample, since non-voters generally did not have the opportunity to be recruited as triplers. Because the Google Sheet was primarily only used for polling place tripling on 3/14/20, Columns (3) and (4) only show results for those who are recorded by Lorain County to have voted that day, who indeed represent the majority of those shared on the spreadsheet (70%). However, there appears to be some inconsistency in recording these dates, so it is worth considering both pairs of regressions.

The main takeaway from these regressions is what they do not show: in contrast to most previous implementations of vote tripling, triplers are not disproportionately higher-propensity voters. In previous studies, higher-propensity voters are overwhelmingly more likely to opt in to vote tripling, much as higher-propensity voters are more likely to undertake all manner of campaign volunteering. That pattern is missing from this data; on the contrary, more frequent voters are slightly less likely to have been triplers. One should interpret the significant coefficients with caution, given the large number of hypotheses being tested; but we can say with some confidence that this ask

[10] These figures are based on estimates after the intervention, as data was not tracked in real time.

[11] Students generally arrived by the carload, and often the spreadsheet was shared to only one student in the car, who then circulated it to others on the drive back. As a result, there are fewer students shared on the spreadsheet than there were triplers.

[12] The intentionally conservative construction of the sample frame, with only recently-registered or November-2019-voting students included in the universe, meant that many Oberlin voters were not included; as a result, it is unsurprising that not all triplers are found in the sample frame. Additionally, some may not match because of name spellings or other issues, although we did screen for common nicknames.

was not exclusively appealing to the most activist of voters. In a sense, the lack of any such pattern is not surprising, given the high pledge rates across the board. Given that around 75% of voters agreed to vote triple (which is significantly higher than, but still roughly commensurate with previous pilots of this model), there simply is not that much headroom to be skimming only the highest-propensity voters off the top of the propensity distribution.

Table 2: Predicting who becomes a tripler

	All who voted (1)	All who voted (2)	All who voted on 3/14/20 (3)	All who voted on 3/14/20 (4)
Partisan score	-.137* (.083)		-.062 (.107)	
Presidential turnout score	-.002 (.023)		-.002 (.032)	
Midterm turnout score	-0.011 (.010)		-.011 (.018)	
Primary turnout score	.019 (.047)		-.048 (.056)	
Voted 2019 general		-.278 (.269)		.237 (.354)
Voted 2018 general		-.581* (.303)		-.973** (.419)
Voted 2017 general		-.042 (.581)		-.386 (.890)
Voted 2016 general		-.681 (.445)		-.388 (.582)
Female	.027 (.275)	-.097 (.252)	-.392 (.338)	-.553 (.340)
Minority	-1.25* (.745)	-1.29* (.744)	-2.26** (1.06)	-2.28** (1.07)
Age	.077 (.129)	.121 (.150)	.207 (.160)	.351* (.194)
Constant	10.29 (6.43)	-3.50 (2.99)	2.67 (8.48)	-6.79* (3.83)
Pseudo R^2	.027	.028	.059	.067
N	490	490	175	175

Standard errors in parentheses. * = p<.1 ** = p<.05 *** = p<.01. As above, voter turnout in prior elections does not correct for whether the individual was eligible or registered to vote in that previous election. Note that the estimate on "female" is somewhat unreliable, as 17% of the sample has gender unknown.

There is also some evidence that voters of color were significantly less likely to opt in to being triplers. But, given that Targetsmart classifies only 7% of March early voters as people of color, this result should be interpreted with some caution.

5. Experimental results

5.1 Main results

For our principal turnout results, we use early vote data provided by Lorain County, which, due to the unique nature of this election, contains all votes in the Ohio primary. Specifically, we downloaded a list of all primary voters as of May 9, 2020, and discarded those flagged as invalid.[13]

Table 3 shows the headline results based on overall turnout, including ballots submitted by mail throughout April. Column (1) shows the primary specification, without any additional controls. As noted above, there are some imbalances across treatment and control, which means it is prudent to add controls — but these controls are only available among the 988 observations in the Targetsmart match. Column (2) shows the headline result only among those observations; Columns (3) and (4) show the main specification with different sets of controls. Overall, the four specifications tell a similar story: treatment likely increased turnout among the treatment group by around 7-8pp. Despite the small sample size, this result is statistically significant, with p-values around .04-.07 in specifications (1), (3), and (4). In concrete terms, turnout in the control group was 47% compared to nearly 55% in the treatment group. The Column (1) estimate is equivalent to generating 70 net votes in the treatment group.

Table 3: Vote tripling mobilization effects

	(1)	(2)	(3)	(4)
Assigned to treatment	.078** (.039)	.062 (.041)	.076* (.040)	.072* (.041)
Rate in control	.47	.489	.489	.489
Controls	None	None	Vote history and demographics	Targetsmart scores and demographics
Sample	All	Tsmart match only	Tsmart match only	Tsmart match only
N	1100	988	988	988

Ordinary Least Squares, OLS, regressions with turnout as dependent variable. Standard errors in parentheses. * = p<.1, ** = p<.05, *** = p<.01.

[13] The intentionally conservative construction of the sample frame, with only recently-registered or November-2019-voting students included in the universe, meant that many Oberlin voters were not included; as a result, it is unsurprising that not all triplers are found in the sample frame. Additionally, some may not match because of name spellings or other issues, although we did screen for common nicknames.

It is somewhat unclear whether the appropriate dependent variable for this analysis is all voting, or early voting during the March 12-16 period. On one hand, we are interested in generating net votes, rather than simply moving votes earlier, which supports measuring all votes, including those that occurred after March 16. On the other hand, the odd circumstances of the Ohio primary — in which in-person election day voting was suddenly canceled, and the deadline to submit mail ballots extended six weeks — are hardly generalizable. Under any normal circumstances, most student voters would not have had a second chance to vote, had they missed the early voting window available to them; the clock would have run out during March early voting. Moreover, no normal GOTV tactic would stop organizing six weeks before an electoral deadline: the unique circumstances allowed results to dissipate slightly throughout the month of April. Arguably, then, what we are most interested in are results in March 12-16 early voting only. These are shown in Table 4, and are, unsurprisingly, somewhat higher than in Table 3. The headline result in Column (1) suggests the intervention produced 82 net votes.

Table 4: Headline results — March 12-16 only

	(1)	(2)	(3)	(4)
Assigned to treatment	.091** (.039)	.080* (.041)	.093** (.040)	.089** (.041)
Rate in control	.440	.456	.456	.456
Controls	None	None	Vote history and demographics	Targetsmart scores and demographics
Sample	All	Tsmart match only	Tsmart match only	Tsmart match only
N	1100	988	988	988

OLS regressions with turnout as dependent variable. Those who voted after March 16 are coded as not voted. Standard errors in parentheses. * = p<.1, ** = p<.05, *** = p<.01.

As noted in Table 1, about 10% of the sample lived off-campus. Because off-campus students were not forced to leave their Oberlin residence on March 16, they faced a different set of incentives than their on-campus counterparts: for off-campus students, regular election day voting appeared to be a viable option (until the cancelation occurred). For these students, early voting was not de facto election day. As such, the more generalizable specification might be the March early voting effect among on-campus students only, for whom early voting *was* essentially equivalent to election day voting.

Table 5 shows that the effect among on-campus students is larger than off-campus students, although, given the small number of off-campus students, the difference in treatment effect (interaction term) is not significant. The effect on on-campus students is incrementally higher, around 9-10pp.

Table 5: Headline results — March 12-16 early voting only — heterogeneity by on/off campus

	(1)	(2)	(3)	(4)
Assigned to treatment	.102** (.041)	.092** (.042)	.094** (.042)	.100** (.042)
Treatment x Off-campus	-.044 (.132)	-.037 (.135)	.006 (.131)	-.046 (.133)
Off-campus	-.300** (.121)	-.308** (.124)	-.397 (.124)	-.331** (.127)
Controls	None	None	Vote history and demographics	Targetsmart scores and demographics
Sample	All	Tsmart match only	Tsmart match only	Tsmart match only
N	1100	988	988	988

OLS regressions with turnout as dependent variable. Those who voted after March 16 are coded as not voted. Standard errors in parentheses. * = p<.1, ** = p<.05, *** = p<.01.

It is worth noting that most of these estimates are probably slight underestimates due to treatment leakages. While triplers were encouraged only to contact people on the list, surely some ended up contacting others, either because they contacted group text threads that included other voters, or because the intervention put them in mind of sending voting reminders, and they ended up being encouraged to do so more broadly. There were several documented examples of this phenomenon, in which reminders were sent to, e.g., the entire track team's text list, without regard, of course, for the treatment assignment of the entire track team.

5.2 Heterogeneity results

Table 6 presents the results of a heterogeneity analysis along several dimensions. Each cell represents the interaction term between treatment and the axis of heterogeneity, among the 988 observations with Targetsmart data. The main finding here is the very large estimate for female voters, which suggests that the entire study's treatment effect is driven by treatment effects[14] among female voters. While the magnitude of this estimate is unrealistic and likely driven by the small sample size, the fact of a significant interaction is nearly indisputable. Friend-to-friend reminders, in this instance, either end up disproportionately targeting female voters, or being disproportionately powerful on them. Because prior Vote Tripling research suggests that triplers disproportionately contact voters of the same gender, and because Section 4 shows that triplers are not disproportionately female, it seems likely that the latter mechanism is operative: in this experiment, friend-to-friend reminders were more effective when targeted at female voters.

Table 6: Heterogeneity results

	Without controls (1)	With controls (2)
Pres. primary turnout score	-.012 (.010)	-.016 (.009)
Ever voted	-.016 (.089)	-.019 (.087)
Voted Nov 2019	-.025 (.086)	-.047 (.085)
Female	.298*** (.095)	.274*** (.092)
Minority	.227 (.159)	.227 (.155)
Age	-.024 (.025)	-.027 (.024)

Each cell presents the coefficient of an interaction term between treatment status and the theorized axis of heterogeneity — each cell is from a separate regression. The full regression in each case is of voter turnout on treatment status, heterogeneity variable, the interaction, and — in column 2 — controls. N for each regression is 988, except for female interaction, where "unknown" gender is dropped, leaving 825 observations. OLS regressions, standard errors in parentheses. * = p<.1, ** = p<.05, *** = p<.01.

There is additionally some weak evidence that the treatment was slightly stronger for those who were less likely to vote, though it is very likely this is due only to chance. It is

[14] Taken at face value, the estimate suggests that women in general vote at lower rates, and that treatment has if anything slightly negative impact on male voters, but the enormous treatment effect on women is enough to make up for both of these effects.

also possible that minorities were more likely to be mobilized by the treatment, but, again we cannot reject that these results are due only to chance.[15]

5.3 Temporal analysis

Because early voting results are recorded day by day, it is possible to trace the effect of treatment on each day of early voting. We expect to see no treatment effect on 3/12, when the lists had not yet been disseminated; limited treatment effect on 3/13, when mobilization activities were limited; and larger effects on 3/14, 3/15, and 3/16, when PPVT was in full swing.

Surprisingly, in Table 7, we see something very different: the implied treatment effect on 3/12 and 3/13 is in fact very large. Given how little intervention had been performed those days, this result is almost surely noise.

Table 7: Temporal analysis

	Control votes	Treatment votes	Percentage point treatment effect	Cumulative pp effect	Total pp effect if total program looked like this day
3/12	2	18	+1pp	+1pp	+47pp
3/13	34	205	+5.8pp	+6.8pp	+16.0pp
3/14	34	163	+1.1pp	+7.9pp	+3.1pp
3/15	11	63	+1.5pp	+9.4pp	+12.8pp
3/16	6	27	+0pp	+9.4pp	+0pp
After 3/16	7	17	-1.6pp	+7.8pp	-21.6pp

[15] If it were true, this smaller treatment effect on minorities would be consistent with the Section 4 finding that fewer minorities signed up to be triplers, given existing evidence that triplers choose targets demographically similar to themselves.

Table 8: Results dropping 3/12 and 3/13

	All voting (1)	Only March early voting (2)	All voting (3)	Only March early voting (4)
Assigned to treatment	.045 (.042)	.060 (.042)	.046 (.044)	.066 (.044)
Rate in control	.354	.317	.394	.370
Controls	None	None	Vote history and demographics	Vote history and demographics
Sample	Dropping 3/12 and 3/13 voters	Dropping 3/12 and 3/13 voters	Dropping 3/12 and 3/13 voters, Tsmart match only	Dropping 3/12 and 3/13 voters, Tsmart match only
N	841	841	751	751

OLS regressions with turnout as dependent variable. In Columns (2) and (4), those who voted after March 16 are coded as not voted. Standard errors in parentheses. * = p<.1, ** = p<.05, *** = p<.01.

For robustness, Table 8 shows what happens if we remove 3/12 and 3/13 from the experiment entirely. Unsurprisingly, there is not enough sample size left to drive significant results, but the estimated effect sizes are substantively still quite large, at 4.5-6.5 percentage points. The p-values are around .28-.30 for columns (1) and (3), and .13-.15 for columns (2) and (4). The temporal analysis suggests the headline point estimate of 7-10pp is probably biased upwards, but the balance of the evidence still implies there was a substantial treatment effect.

5.4 Impact on a per unit basis

The estimates presented in this paper are not immediately comparable to other RO ITT results since individual treatment voters were likely to have been contacted more than once, by multiple triplers. Table 9 attempts to put these results in more consistent context by showing the impact in terms of units a campaign can easily account for: impact per voter contact, or impact per tripler recruited. The results are based on a total effect of 70 net votes, based on Column (1) of Table 3. (If we use the results from Table 8, the figure is 30-40 net votes.) Of course, these per-contact results rely on several assumptions: how many triplers actually sent texts, how many texts they sent, and how much of the net effect came from non-PPVT organizing. The table shows effects under multiple sets of assumptions as to the values of these parameters.

Table 9: Net votes on a per contact basis, under various assumptions

	Assuming...			Then:	
	Number of triplers who sent reminders was: (1)	...and the avg number of reminders sent was: (2)	...and the portion of the effect via non-PPVT list-sharing was: (3)	The avg impact per recruited tripler is: (4)	And the avg impact per friend-to-friend contact is: (5)
1	325	5	.2	.172	.034
2	325	5	.4	.129	.026
3	325	2	.2	.172	.086
4	325	2	.4	.129	.065
5	200	5	.2	.280	.056
6	200	5	.4	.210	.042
7	200	2	.2	.280	.140
8	200	2	.4	.210	.105

While the results are of course subject to significant uncertainty, it seems reasonable to conclude that a single tripler recruited by PPVT produced around .1-.3 net votes — or .05-.15 net votes, if we accept the smaller headline result in Table 8. Existing research on relational organizing suggests that a single friend-to-friend contact is worth around .04-.1 net votes,[19] which is roughly consistent with most of these models.

Two other measures of impact are subject to less uncertainty:

- Net votes per hour of PPVT. The implementation included 15 hours of PPVT. Again assuming 70 net votes, 42-56 generated by PPVT, PPVT generates 2.8-3.7 net votes per hour. If we use the conservative results from Section 5.3, the figure is 1.4-1.9 net votes per hour.
- Cost per vote. The implementation cost a total of $35 in materials, equating to a trivial 50 cents per net vote. The implementation further required 21 person-hours of — uncompensated — organizer time. If an organizer were paid $20/hour, the entire project would have cost $6.50 per net vote. If we use the conservative results from Section 5.3, the figure is around $13 per net vote.

6. Discussion

The 7-10pp ITT turnout effect estimated in this study is very large by any measure. Due to the specific nature of the intervention, its interpretation is not exactly straightforward, but we suggest there are a few simple ways of understanding this effect, if it is to be taken at face value:

- An organizer doing PPVT can generate around 30 net votes per day[16]. If we use the conservative results from Section 5.3, the figure is around 15 net votes per day.
- Each vote tripler recruited during PPVT is worth .1-.3 net votes. With the more conservative results, .05-.15 net votes.
- PPVT and list-based RO on college campuses can generate votes at a rate of $6.50 per vote. With the more conservative results, $13 per vote.

There are two notable concerns about the internal validity of the study.

First, the dependent variable is, largely, early voting only. While we run the headline results both with and without all votes cast through April, the cancellation of the originally planned election and the rapidly changing political circumstances mean that we are, in a practical sense, still looking only at early voting; election day voting, essentially, did not occur. This is a red flag, since past turnout work has shown that interventions encouraging early voting have often managed not to generate net new votes, but simply move the same votes earlier in the process. Interventions produce earlier votes, but not net votes.

But the heterogeneity of effect between on- and off-campus students argues against this conclusion. The concern is that an intervention promoting early voting does not change how many people vote overall, but merely moves votes from T_2(election day) to T_1(early vote). If this is true, then populations who have the opportunity to vote during either T_1 or T_2 will have larger apparent effect sizes than populations that do not. That is, a population with both voting options available will have a fictitious treatment effect in T_1, comprised of people who would otherwise have voted in T_2; but a population without T_2 available will vote in T_1 at the expected T_{1+2}rates, and the illusory mechanism of moving votes from $T_2$to T_1 will be absent. If this theory held, then off-campus students would see a larger effect during the early vote period. But, as shown in Section 5, they do not; the effect, if anything, appears to be significantly larger on on-campus students. This finding

[16] Based on an unpublished meta-analysis of the TOT impact of relational contacts conducted by Vote Tripling.

provides additional assurance that the results are not spurious findings due to the temporal moving of votes.

Second, as noted in the results section, the estimates may be slightly biased downwards across the board because of likely spillover to the control group. As such, we suggest the actual results reported in this study are a slight underestimate of the actual effect.

The fact that this experiment was implemented during a primary provides intriguing evidence that Vote Tripling and other relational tactics can be reliably used in primaries, despite the fact that, intuitively, it would seem primary voters have less reliable information on their friends' allegiances in primaries than in generals. However, we caution that Oberlin in March 2020 was a bit of a special case. 78% of the sample universe in this study was registered in precincts Oberlin City #7 and Oberlin City #8, which voted respectively 83% and 86% for Bernie Sanders.[17] Anecdotally, those percentages may have been still higher among Oberlin's student body itself. As a result, in this special case, voters could have been reasonably confident of their friends' affiliations, something which may not be true of all communities in all primaries.

There are a variety of concerns that limit the external validity of this study, more broadly. These are presented below and summarized in Table 10.

First, college campuses are arguably more tight-knit communities than average, meaning that it is easier to create a list of voters that other voters will know. The difficulty of creating lists obviously has implications for translating this intervention to other contexts. The obvious implications are on measurement: in another community, where voters did not know anyone on the list they were shown, it is far less likely we would see turnout lift on voters placed on the list. However, it is ambiguous whether the issue would have programmatic impact. Absent a list, organizers could pursue a more traditional PPVT approach, and ask voters to think of three friends to remind. Does working without a list this way cause fewer voters take action, and/or does it make those actions less effective? Circumstantial evidence from the implementation suggests it may, because the list of voters was an appealing piece of information for triplers: highlighting which friends had not made it to the polls, and ensuring that any reminders would not be wasted — condescendingly — on those who had already voted. But it is hard to answer this question rigorously, and nearly impossible to design an experiment that would answer it, given the endogeneity of the list to the measurement itself.

[17] Results are available from the Lorain County website: https://49d11d30-6b21-4c94-868b-77f4f0c1339c.filesusr.com/ugd/2568d0_b3ca7b1e5e034e0a832f685b1 fae637d.pdf, accessed from http://www.loraincountyelections.us/2020-primary-election.

Second, the early voting context meant it was possible to regularly update the list by removing those who had already voted, something that is less feasible in an election day PPVT implementation, even if it were possible to identify a list of voters known to potential triplers. Lists containing all voters, rather than only those who have not yet voted, are — again circumstantially — probably less appealing to potential triplers, and using such an uncurated list may lead to worse results.

Third, again because college campuses are tight-knit communities, students are arguably more willing to make contacts to fellow students than average voters would be to the general public. This effect is likely compounded by the fact that younger voters have larger social networks in general. It is possible that, in a non-college context, fewer voters would pledge to triple, fewer would follow through, and the contacts would be less effective. This is especially true given the unique context of the experiment, whose extraordinary circumstances may have engendered a feeling of communal spirit, and a desire to help everyone in the community vote, that may not exist in normal times, or in less tight-knit communities. This aspect would cause these results to skew positive, at least insofar as similar interventions are attempted in non-college contexts.

Fourth, the early voting location being so far from campus, and the circumstances being so extraordinary, would have generated a non-standard pool of voters at the polls: Anyone who made it to vote was, by definition, a fairly committed voter, which may have generated especially fertile territory to recruit vote triplers. As noted in Section 4, voter propensity did not predict tripler sign-up among those who voted; but the selection may already have been done by who showed up to vote at all. It is possible that, in a normal polling place, the pool of voters would contain fewer, or less effective, triplers.

Fifth, the extraordinary circumstances and the need to carpool to the early voting location meant that students had pertinent information to give their friends, which may have provided a useful pretext for communication that may have been perceived as condescending in normal times. While campaigns may be able to replicate this effect in 2020 through the promotion of mail voting, it may be hard to mimic the highly unique situation of this study.

Sixth, the fact that most students carpooled to the polling location and had a 20-minute ride home awaiting them perhaps gave them an excellent opportunity to send reminders in the car, when under normal circumstances they might have been distracted before doing so. The set-up, that is, may have increased the follow-through rate on a tripling pledge.

Seventh, this intervention was implemented over the course of several days of early voting, mostly not on the last day. While this may not be an issue in November 2020, when much voting may shift early to promote social distancing: PPVT has traditionally been implemented on election day, when it is the last day to vote, and so a potential voter must act on a reminder they receive very quickly. The fact that voters could have reacted to the reminder by making a plan for the following day may have inflated this effect relative to what we would see doing the same intervention on election day. On the other hand, the lack of immediate urgency may also increase procrastination, so it is not obvious which direction this effect runs.

Eighth, the fact that the polling place was so far and the circumstances so dire means that voting was effectively harder for this treatment group than it usually is for the average voter. As a result, it would have been arguably harder to convince students to vote than it would normally be, and so each contact might have had less effect than it would have on a normal election day, with the polling place conveniently located and no time-pressing need to move out of a dormitory.

These considerations are summarized in Table 10. All told, given that the balance of these considerations serve to inflate the effect size, it seems safe to say that the large effect seen here was closer to an upper bound. However, even if the true effect were only 25% that estimated here, the intervention's cost at $20/organizer-hour would still be only $26 per net vote. The costs of course could be substantially lowered if volunteers were involved.

Table 10: Biases limiting external validity

	Factor	Direction of bias
1	Tight-knit community makes list creation easy	Ambiguous, probably pos.
2	Early voting makes lists more valuable, as those who already voted can be reliably removed	Ambiguous, probably pos.
3	Tight-knit community means people will be more willing to make reminders to friends	Positive
4	Difficulty of voting means those who did vote are especially committed	Positive
5	Difficulty of voting means friends had pertinent information to transmit to friends, providing useful pretext for reminder communication	Positive
6	Long drive home with nothing to do but send reminders	Positive
7	Reminders were being sent throughout several days of early voting, and voters could have then acted on the reminder the following day	Probably positive
8	Early voting at a faraway location was onerous for those receiving reminders, even if they did want to get out to the polls	Negative

A major unanswered question in scaling up the learnings of this study is determining the importance of list-browsing in recruiting and converting triplers. If lists are critical to PPVT's success, then much work remains to be done in designing lists on which voters are likely to recognize friends and family: in most non-university contexts, depending on social geography, this may be a non-trivial challenge. If they are not, if PPVT is equally powerful when voters are asked to simply think of three friends to remind, then this could be scaled to every targeted polling place in the country.