Online Appendix
Immigrant Integration in the United States: The Role of Adult English Language Training

Blake H. Heller and Kirsten Slungaard Mumma

June 1, 2023

A Missing Date of Birth, Level, or Class Time Preference

In Appendix Table A8, we conduct a series of bounding exercises to determine whether covariates or missing data affect our main results. Column (1) presents our main results, for comparison. Column (2) presents estimates from a model that omits individual-level controls. Results are substantively the same.

In column (3), we consider whether missing DOB data biases our results. We are missing DOB for 2.8% of all individuals who we observe as first-time lottery applicants between fall 2008 and spring 2016. (See Appendix Table A7 for information on missing data). Since DOB is required to match to outcome data, these individuals are dropped from our sample in our main results. As a sensitivity test, we impute favorable outcomes for lottery non-winners with missing DOB and unfavorable outcomes for lottery winners' observations with missing DOB. Specifically, we impute that treatment observations with missing DOB never register to vote, and that control observations have a 20% rate of voting and voter registration. For reported earnings outcomes, we impute that all treatment observations with missing DOB data did not report any earnings, but that control observations with missing DOB data reported earnings at the median of the distribution of a given earnings outcome for the sample of control individuals with a positive value for that outcome (e.g., the 50th percentile of the control group earnings distribution for year three reported earnings, conditional on having positive reported earnings in year three). We present estimates using these assumptions in Column (3) of Appendix Table A7. Estimates are statistically indistinguishable from our main results.

We are missing baseline English proficiency level for 2.9% of individuals. Since baseline English level is required to identify an individual's lottery group, we drop these observations from our main results. In columns (4) and (5) of Appendix Table A8, we estimate treatment effects under the assumption that all of these observations are beginners (Column 4) or advanced (Column 5) students. Classifying applicants with missing levels as beginners, the most common observed category, increases our estimated effects. Classifying all applicants with missing level data as advanced students yields estimates that are indistinguishable from our main results.

We are missing time availability data for 15.1% of our analytic sample. In our main results, we classify these applicants as participating in evening lotteries, since the ratio of evening to morning applications is over four to one for individuals with known preferences. Estimates in column (5) show that our results are not sensitive to whether or not we include individuals with missing availability. In column (7), we consider an alternative test where we impute morning availability to individuals with unknown time preferences instead of evening availability. Again, results are similar.

Finally, since we match to outcome data using every combination of name and DOB observed in our three administrative datasets, we consider whether incidences of name and DOB differ by lottery outcome. In our analytic sample, we find no statistically significant differences in incidences of names or dates of birth by lottery outcome. (See results in Appendix Table A7).

B Out-of-State Mobility

If winning access to the FAESL+ program impacts the probability an individual remains in the state—for instance, by creating stronger ties to the local community—inter-state migration could bias our results since outcome data are only measured in the state of Massachusetts. In Appendix Table A9, we assess this possibility in three ways. First, we obtain voting records from public sources for four of the top six destination states of Massachusetts residents who move within the United States, including three of the five states that share a border with Massachusetts (Rhode Island, Connecticut, and New York) and Florida (U.S. Census Bureau, 2018b).[1] We match individuals to these records using name and DOB.[2] We match 26 lottery winners and 85 lottery non-winners to out-of-state voting records. In column (1) of Appendix Table A9, we test whether winning an ESOL enrollment lottery predicts being a registered voter in any of these four destination states. We find no evidence that winning the lottery is related to out-of-state voter registration.[3]

Second, we consider whether we find evidence of differences in *intra*-state migration. In columns (2) and (3), we estimate the effect of winning a lottery to attend the FAESL+ program on registering to vote or reporting earnings within Massachusetts, but outside of Framingham.[4] Effects are insignificant and point estimates are positive in both cases; if anything, this suggests lottery winners are more likely to appear outside the Framingham area than those who do not.

Finally, we consider whether we find patterns in earnings data that are consistent with differences in out-of-state mobility. In our employer-reported earnings records, we test whether winning the lottery predicts that individuals with stable earning histories (defined as ever reporting earnings for four consecutive quarters) suddenly and permanently stop reporting earnings in a future quarter. In column (4), we show that winning the lottery does not predict that individuals fit this pattern of reported earnings overall. In column (5), we find that winning the lottery does predict this pattern of reported earnings when we restrict our sample to the smaller group of individuals with stable earnings histories. Lottery winners with a stable prior earnings history are 9.9 percentage points less likely to suddenly stop reporting earnings than non-winners.

While there are many reasons an individual may stop reporting earnings, including a positive effect of ESOL services on stable employment, we conduct a series of robustness checks to bound the influence of possible out-of-state mobility on our estimates and present these in Appendix Table A10. First, we test whether the difference in rates of attrition from the reported earnings data can explain the main results. To do this, we identify the final post-lottery quarter an individual reported earnings for everyone in our sample with stable post-lottery earnings histories and carry that quarter's earnings forward through the end of the panel. This imposes the assumption that all "stopping out" from stable earners is due to out-of-state mobility and that individuals who "stop

[1] Voting records from California and New Hampshire, the two other top destination states, are not readily available to the public.

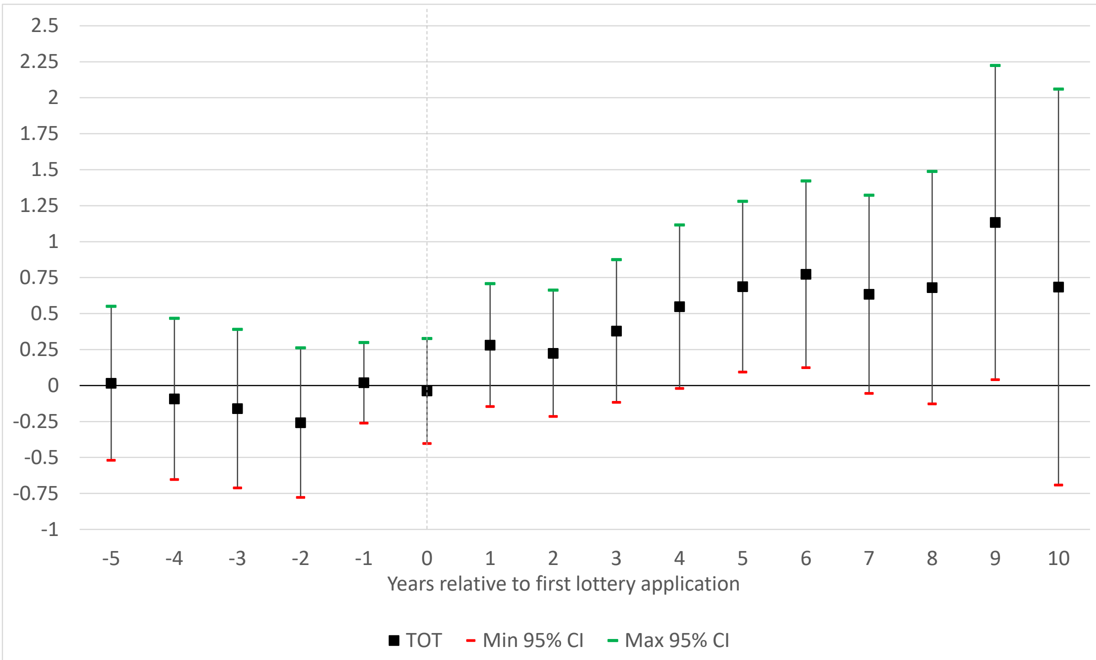
[2] Records include fist name, last name, and DOB for currently registered voters. For Rhode Island records, we use name and year of birth since DOB is not made available in these files. Sources: https://www.connvoters.com/ (CT, accessed May 30, 2020); https://rivoters.com/ (RI, accessed May 30,2020); https://www.elections.ny.gov/FoilRequests.html (NY, received January, 2020); https://flvoters.com/ (FL, accessed August 11, 2020).

[3] To get a rough sense of inter-state migration rates among immigrants living in Massachusetts, ACS records from 2008-2019 show that the number of foreign-born individuals who report having lived in Massachusetts during the prior year and are observed living in another U.S. state ranges from 1.7-2.5% of the total foreign-born ACS sample in Massachusetts in that prior year (Ruggles et al., 2021). Akee and Jones (2019) follow 2005-2007 ACS cohorts in linked tax data and find that just under 40% of recent immigrants in their sample stopped reporting earnings by 2015, which they interpret as a rough estimate of return mobility rates.

[4] We define Framingham as the area including the following five zip codes: 01701, 01702, 01703, 01704, or 01705. Results are similar if we also include voting or earnings in zip codes of all cities and towns contiguous to Framingham.

out" would be earning as much as they did before if we were able to observe their out-of-state earnings. Since more control observations suddenly and permanently stop reporting earnings, this imputation affects more control observations, "correcting" for differential attrition. In column (2) of Appendix Table A10, we conduct a more conservative test by carrying forward earnings for our control group only, imposing the assumption that all "stopping out" in the control group is due to out-of-state mobility but all "stopping out" in the treatment group are quarters with no earnings. Under these tests, the estimated effects on reported earnings attenuate by 25-55% but remain statistically indistinguishable from the main results.

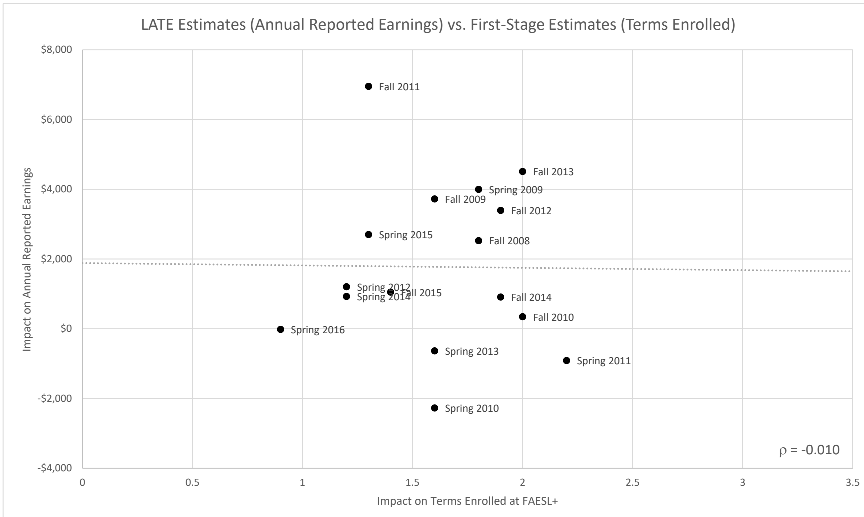
Figure A1: Annual effects on ln(reported earnings), by year since lottery



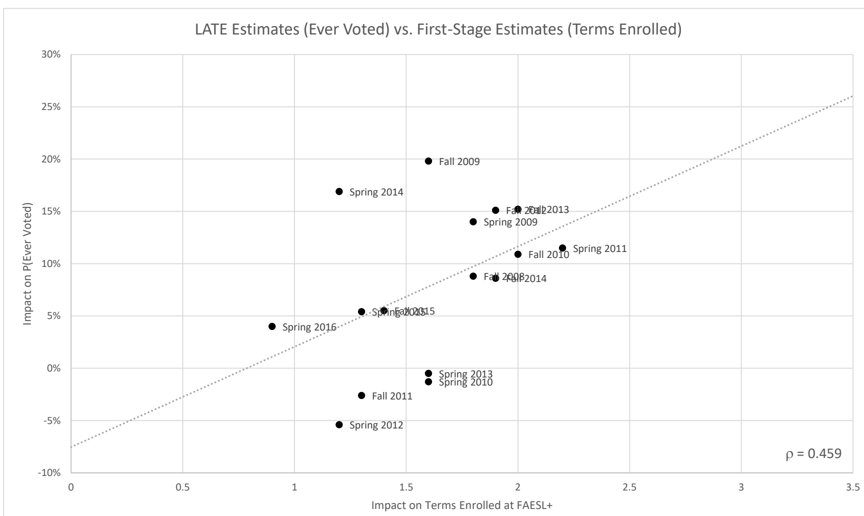
Note: Year of reported earnings is defined relative to first lottery (year=0). LATE point estimates and heteroskedasticity-robust confidence intervals are calculated from 2SLS IV estimates using equation (5) of the effect of enrolling in the FAESL+ program on the natural logarithm of reported income in the indicated year. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. Appendix Table A3 records the point estimates plotted here.

Figure A2: Treatment Effects versus First-Stage Estimates by Lottery Cohort

(a)



(b)



Note: LATE estimates on Annual Reported Earnings (Subfigure A) and P(Ever Voted) (Subfigure B) are plotted on the Y-axis against first-stage estimates of the impact of winning a lottery to attend FAESL+ on terms enrolled on the X-axis. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; and an indicator for missing gender. LATE estimates in Subfigure A cluster standard errors at the individual level and control for baseline quarterly earnings as well as period fixed effects. Appendix Tables A1 (column 3), A14 (column 2) and A15 (column 3) record the point estimates plotted here.

Table A1: Lottery Balance and First-Stage Estimates by Semester

	P-Value from Joint F-Test (1)	First-Stage (Ever Enrolled) (2)	First-Stage (# of terms) (3)
Fall 2008	0.119	0.525	1.769
Spring 2009	0.735	0.508	1.778
Fall 2009	0.156	0.333	1.550
Spring 2010	0.684	0.551	1.562
Fall 2010	0.063	0.419	1.971
Spring 2011	0.015	0.528	2.237
Fall 2011	0.797	0.457	1.291
Spring 2012	0.842	0.396	1.231
Fall 2012	0.010	0.489	1.878
Spring 2013	0.311	0.466	1.559
Fall 2013	0.219	0.385	2.027
Spring 2014	0.161	0.486	1.210
Fall 2014	0.707	0.554	1.871
Spring 2015	0.350	0.604	1.299
Fall 2015	0.316	0.733	1.447
Spring 2016	0.900	0.926	0.931

Note: Column (1) reports the p-value from a joint test of the significance of differences between treatment and control group means of all covariates in Panel A of Table 2 for each semester's lottery. In individual lotteries, imputed race characteristics are included for all race and ethnicity subgroups with at least five observations. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. Baseline earnings are available and included in balance tests beginning in fall 2010. Columns (2) and (3) report first-stage estimates from regressions of an indicator for ever enrolling at FAESL+ (Column 2) or the number of terms an individual enrolled FAESL+ (Column 3) on an indicator for being offered a seat at FAESL+ on an individual's first lottery attempt, controlling for all covariates and lottery fixed effects described above. Standard errors are omitted for first-stage estimates. All p-values are < 0.001.

Table A2: Effect on Probability of Having Registered to Vote by Year Since Lottery

Year (1)	Control Mean (2)	Ever Enrolled (3)
-5	0.01	-0.008
		(0.006)
-4	0.01	-0.004
		(0.007)
-3	0.01	0.001
		(0.008)
-2	0.01	-0.002
		(0.008)
-1	0.01	-0.001
		(0.010)
0	0.02	-0.003
		(0.012)
1	0.03	0.002
		(0.013)
2	0.03	0.014
		(0.015)
3	0.04	0.024
		(0.017)
4	0.05	0.051**
		(0.019)
5	0.06	0.058**
		(0.020)
6	0.06	0.071**
		(0.021)
7	0.07	0.079**
		(0.022)
8	0.08	0.085**
		(0.023)
9	0.08	0.090**
		(0.024)
10	0.08	0.089**
		(0.024)

Note: Year is defined relative to first lottery (year=0). Column (1) reports the proportion who had registered to vote by the indicated year among individuals in our sample who did not win their first lottery attempt. Column (2) reports 2SLS IV estimates using equation (3) of the effect of enrolling in the FAESL+ program on having registered to vote by the indicated period, with heteroskedasticity-robust standard errors in parentheses. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. The coefficients reported here are plotted in Figure 2. N=4,761. $^{*} = p < 0.05$, $^{**}p < 0.01$.

Table A3: Effects on Annual earnings by Year Since Lottery

Year	Control Mean (1)	Earnings $ (2)	Earnings Ln($) (3)	Observations (4)
-5	$949	123	0.016	1,022
		(603)	(0.273)	
-4	$1,129	120	-0.093	1,480
		(705)	(0.286)	
-3	$1,424	264	-0.160	1,934
		(716)	(0.281)	
-2	$2,162	-713	-0.258	2,363
		(713)	(0.265)	
-1	$2,826	-189	0.019	2,864
		(359)	(0.143)	
0	$3,549	-478	-0.038	3,597
		(459)	(0.186)	
1	$3,720	467	0.281	4,353
		(635)	(0.218)	
2	$3,765	1,514*	0.224	4,761
		(731)	(0.224)	
3	$4,070	1,822*	0.379	4,345
		(884)	(0.253)	
4	$4,058	2,814**	0.548	3,739
		(1,057)	(0.290)	
5	$4,354	2,791*	0.687*	3,281
		(1,184)	(0.303)	
6	$4,590	2.381	0.773*	2,827
		(1,290)	(0.331)	
7	$4,570	2,173	0.634	2,398
		(1,379)	(0.351)	
8	$4,331	2,449	0.680	1,897
		(1,571)	(0.412)	
9	$3,894	4,981*	1.133*	1,164
		(2,262)	(0.557)	
10	$2,943	2,859	0.684	408
		(2,857)	(0.702)	

Note: Year is defined relative to first lottery (year=0). Column (1) reports mean reported earnings in the indicated year among individuals in our sample who did not win their first lottery attempt. Columns (2) and (3) report 2SLS IV estimates using equation (5) of the effect of enrolling in the FAESL+ program on reported earnings (Column 2) and their natural logarithm (Column 3) in the indicated year, with heteroskedasticity-robust standard errors in parentheses. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. The coefficients reported in column (2) are plotted in Figure 3 and coefficients reported in column (3) are plotted in Appendix Figure A1. $^{*} = p < 0.05$, $^{**}p < 0.01$.

Table A4: Effect on P(Ever reporting earnings within selected ranges)

Earnings Range	Control Mean (1)	LATE (2)
\$0 – \$10,000	0.114	0.011
		(0.025)
\$10,000 – \$20,000	0.103	0.034
		(0.025)
\$20,000 – \$30,000	0.091	0.060**
		(0.023)
\$30,000 – \$40,000	0.077	0.016
		(0.020)
\$40,000 – \$50,000	0.046	0.014
		(0.016)
\$50,000 – \$60,000	0.030	0.021
		(0.013)
\$60,000 – \$70,000	0.014	0.029**
		(0.011)
\$70,000 – \$80,000	0.009	0.006
		(0.008)
\$80,000 – \$90,000	0.003	0.004
		(0.005)
\$90,000 – \$100,000	0.003	0.004
		(0.004)
Over \$100,000	0.003	-0.0001
		(0.004)
Observations		4,761

Note: Column (1) reports the proportion who ever reported annual earnings in the indicated range among individuals in our sample who did not win their first lottery attempt. Column (2) reports 2SLS IV estimates using equation (3) of the effect of enrolling in the FAESL+ program on reporting earnings in the indicated range, with heteroskedasticity-robust standard errors in parentheses. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. The coefficients reported here are plotted in Figure 4. $^{*} = p < 0.05$, $^{**}p < 0.01$.

Table A5: Alternative Specifications, Effects on Annual Employer-Reported Earnings

	Control Mean (1)	Ever Enrolled [Earnings in $] (2)	Ever Enrolled [Ln(Earnings in $)] (3)
A. Balanced Panels			
Annual Reported Earnings, through Y_1	$3,687	-150	0.072
		(529)	(0.192)
		7,194	7,194
Annual Reported Earnings, Y_2–Y_5	$4,002	2,131*	0.520
		(960)	(0.278)
		13,124	13,124
Annual Reported Earnings, Y_6–Y_9	$3,694	3,948*	0.870
		(1,961)	(0.532)
Observations		4,656	4,656
B. Reweighted Estimates ($W_p = 1/N_p$)			
Annual Reported Earnings, through Y_{10}	$3,989	2,240*	0.570*
		(935)	(0.258)
		32,770	32,770
Annual Reported Earnings, Y_2–Y_{10}	$4,071	2,692*	0.653*
		(1,086)	(0.294)
		24,820	24,820

Note: Results in Panel A are estimated in balanced panels where the sample is restricted to individuals whose reported earnings over the range of post-lottery years indicated in each row could be observed in reported earnings data from 2010–2019. In Panel B, an unbalanced panel is used to generate reweighted estimates where observations are weighted by the inverse of the number of observations in the sample in a given earnings year, where an observation's year is defined relative to the date of an individual's first lottery application (year=0). Column (1) presents the mean of each outcome for individuals in our sample who did not win their first lottery attempt (weighted as described above for Panel A). All outcomes defined over post-lottery periods only. Columns (2) and (3) present 2SLS IV estimates of the impact of ever enrolling at FAESL+ on the outcomes listed in each row, with heteroskedasticity-robust standard errors in parentheses followed by the number of observations that contribute to each estimate. Estimates in each panel are calculated by equation (5) using a longitudinal dataset of applicant-by-year observations (unbalanced panel), with standard-errors clustered at the individual level, with outcomes measured in unadjusted dollars (Column 2) or their natural logarithm plus $1 (Column 3). All estimates include covariates, period fixed effects, and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. $^* = p < 0.05$.

Table A6: Placebo Tests

	Sample (1)	Control Mean (2)	Ever Enrolled (3)
A. Pre-Lottery Voting and Voter Registration			
Registered to Vote	F2008-S2016	0.01	-0.001
			(0.011)
			4,761
Voted	F2008-S2016	0.01	0.002
			(0.010)
Observations			4,761
B. Pre-Lottery Matched to Earnings			
Ever Matched	F2010-S2016	0.15	0.022
			(0.027)
			3,174
Quarters Matched	F2010-S2016	1.04	0.206
			(0.251)
Observations			3,174
C. Pre-Lottery Average Annual Earnings			
Annual Earnings, through Y_{-5}	S2011-S2016	$1,900	-106
			(723)
Observations			9,663

Note: All outcomes defined over pre-lottery periods only. Column (2) presents the mean of each pre-lottery outcome for individuals in the analysis sample indicated in column (1) who did not win their first lottery attempt. Column (3) presents 2SLS IV estimates assessing whether ever enrolling at FAESL+ predicts the pre-lottery outcomes listed in each row, with heteroskedasticity-robust standard errors in parentheses followed by the number of observations that contribute to each estimate. Results in Panels A and B are estimated using equation (3) in a dataset that is unique at the individual-level. Results in Panel C are estimated using equation (5) in a longitudinal dataset that is unique at the individual-by-year level, with standard errors clustered at the individual level. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; and an indicator for missing gender. Panels B and C add baseline earnings as a covariate. Panel C adds period fixed effects. Descriptions of placebo tests are presented in the section IV.D of the text. $^{*} = p < 0.05$, $^{**} = p < 0.01$.

Table A7: Missingness and Incidence of Names and Date of Birth

	All (1)	Won (2)	Did not win (3)	t-stat (p-value) (4)	Observations (5)
Missing DOB	0.028	0.017	0.032	2.88	5,031
				(0.004)	
Missing Level	0.029	0.100	0.002	19.02	5,031
				(0.000)	
Unknown Availability	0.151	0.025	0.196	14.85	4,761
				(0.000)	
Variations of First Names	1.613	1.595	1.620	0.75	4,761
				(0.454)	
Variations of Surnames	2.530	2.544	2.525	0.23	4,761
				(0.816)	
Variations of DOBs	1.060	1.061	1.059	0.17	4,761
				(0.862)	

Note: Missing DOB and Missing Level samples include all individuals who applied to the FAESL+ program for the first-time between fall 2008 and spring 2016. An incident of a name or date-of-birth is defined as a unique iteration of that name or date-of-birth as observed in an administrative dataset. These combinations include iterations of first names with and without middle name (e.g., "Oprah Gail" and "Oprah"), iterations of surnames with and without middle name (e.g., "Gail Winfrey" and 'Winfrey"). If we observe an individual with multiple first, last, and/or middle names, we iterate all possible name combinations (e.g., an individual who appears as both "Carlos Irwin Estévez" and "Charlie Sheen" would generate additional observations for "Carlos Sheen," "Carlos Irwin Sheen," "Charlie Irwin Sheen," "Charlie Estévez," and "Charlie Irwin Estévez"). All other samples are limited to individuals in our analytic sample, which is restricted to individuals who applied to FAESL+ for the first-time between fall 2008 and spring 2016 who have non-missing date-of-birth and initial English level information.

Table A8: Robustness Checks

	Main Results (1)	No Covariates (2)	Impute Outcomes (Missing DOB) (3)	Impute Level (Beginner) (4)	Impute Level (Advanced) (5)	Exclude Missing Availability (6)	Impute Availability (AM) (7)
A. Voting and Voter Registration							
Registered to Vote	0.090**	0.094**	0.080**	0.156**	0.104**	0.087**	0.090**
	(0.022)	(0.020)	(0.022)	(0.028)	(0.023)	(0.024)	(0.022)
	4,761	4,761	4,884	4,890	4,890	4,040	4,761
Voted	0.078**	0.083**	0.066**	0.127**	0.086**	0.076**	0.074**
	(0.021)	(0.019)	(0.021)	(0.026)	(0.022)	(0.023)	(0.021)
Observations	4,761	4,761	4,884	4,890	4,890	4,040	4,761
B. Matched to Earnings Data							
Ever Matched	0.042	0.048	0.028	0.095**	0.048	0.032	0.054
	(0.028)	(0.032)	(0.029)	(0.034)	(0.029)	(0.031)	(0.028)
	4,761	4,761	4,884	4,890	4,890	4,040	4,761
Quarters Matched	1.640*	1.836*	1.407*	2.655**	1.810**	1.586*	1.744**
	(0.670)	(0.756)	(0.674)	(0.786)	(0.687)	(0.724)	(0.663)
Observations	4,761	4,761	4,884	4,890	4,890	4,040	4,761
C. Average Annual Earnings							
Annual Earnings, through Y_{10}	1,843*	1,807	1,415	2,517**	1,880*	1,825*	1,844*
	(771)	(939)	(790)	(886)	(778)	(823)	(749)
Observations	32,770	32,770	33,649	33,571	33,571	29,097	32,770
Annual Earnings, Y_2–Y_{10}	2,388**	2,301*	1,891*	3,144**	2,423**	2,459*	2,392**
	(911)	(1029)	(932)	(1035)	(922)	(978)	(884)
Observations	24,820	24,820	25,514	25,370	25,370	22,385	24,820

Note: Results in Panels A and B are estimated using equation (3) in a dataset that is unique at the individual-level, with heteroskedasticity-robust standard errors in parentheses followed by the number of observations that contribute to each estimate. Results in Panel C are estimated using equation (5) in a longitudinal dataset that is unique at the individual-by-year level, with standard errors clustered at the individual level. All outcomes defined over post-lottery periods only. All estimates include lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. All estimates except those in column (2) add covariates, including gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; and an indicator for missing gender. Panels B and C add baseline earnings as a covariate. Panel C adds period fixed effects. Descriptions of each robustness test are presented in the section IV.D of the text. $^* = p < 0.05$, $^{**} = p < 0.01$.

Table A9: Mobility Tests

	Registered to Vote in RI, CT, NY, or FL	Registered to Vote in MA Outside Framingham	Reported Earnings Outside Framingham	Stopped Reporting After 4Q > $0	Stopped Reporting After 4Q > $0
	(1)	(2)	(3)	(4)	(5)
Won Lottery	-0.004	0.014	0.016	-0.014	-0.099**
	(0.005)	(0.008)	(0.014)	(0.009)	(0.036)
Observations	4,761	4,761	4,761	4,761	926
Sample Restriction	None	None	None	None	4Q>$0

Note: Results are estimated using equation (3) in a dataset that is unique at the individual-level, with heteroskedasticity-robust standard errors in parentheses. In columns (2) and (3), "Outside Framingham" is defined by excluding observations from the five zip codes that comprise the city. In columns (4) and (5), the outcome variable, "stopped reporting earnings" is a binary indicator that takes on a value of one for any individual who is never again observed reporting earnings after being observed reporting earnings in any four past consecutive quarters, and zero otherwise. All outcomes defined over post-lottery periods only. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; and an indicator for missing gender. Columns (3), (4), and (5) add baseline quarterly earnings as a covariate. Descriptions of each test are presented in the section IV.D of the text. $^{*} = p < 0.05$, $^{**} = p < 0.01$.

Table A10: Mobility Robustness Checks

	Main Results	Carry Forward Last Earnings: Full Sample with 4Q>\$0	Carry Forward Last Earnings: Control Only with 4Q>\$0
	(1)	(2)	(3)
Average Annual Earnings, through Y_{10}	1,843*	1,394	866
	(771)	(988)	(972)
Observations	32,770	32,770	32,770
Annual Earnings, Y_2–Y_{10}	2,388**	1,778	1,095
	(911)	(1,077)	(1,059)
Observations	24,820	24,820	24,820

Note: Results are estimated using equation (5) in a longitudinal dataset that is unique at the individual-by-year level, with heteroskedasticity-robust standard errors clustered at the individual level. All outcomes defined over post-lottery periods only. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. Descriptions of each test are presented in section IV.D of the text. ($^{*} = p < 0.05$, $^{**} = p < 0.01$).

Table A11: Alternative IV Estimates: Number of Terms as Treatment

	Control Mean (1)	Number of Terms (2)
A. Voting and Voter Registration		
Ever Registered to Vote	0.07	0.028**
		(0.007)
		4,761
Ever Voted	0.06	0.024**
		(0.007)
Observations		4,761
B. Matched to Earnings Data		
Ever Reported Earnings	0.21	0.013
		(0.009)
		4,761
Quarters with Earnings	3.78	0.509*
		(0.209)
Observations		4,761
C. Average Annual Earnings		
Annual Earnings, through Y_{10}	\$4,022	540*
		(228)
		32,770
Annual Earnings, Y_2–Y_{10}	\$4,147	682**
		(263)
Observations		24,820

Note: Column (1) presents the mean of each outcome for individuals in our sample who did not win their first lottery attempt. All outcomes defined over post-lottery periods only. Column (2) presents 2SLS IV estimates of the impact of enrolling at FAESL+ for one term/semester on the outcomes listed in each row, with heteroskedasticity-robust standard errors in parentheses. Results in Panels A and B are estimated using an adaptation of equation (3) that replaces the binary indicator for program attendance with the number of terms an individual attended FAESL+ in a dataset that is unique at the individual-level. Results in Panel C are estimated using an adaptation of equation (5) that replaces the binary indicator for program attendance with the number of terms an individual attended FAESL+ in a longitudinal dataset that is unique at the individual-by-year level. Standard errors in Panel C are clustered at the individual level. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; and an indicator for missing gender. Panels B and C add baseline earnings as a covariate. Panel C adds period fixed effects. $^* = p < 0.05$, $^{**} = p < 0.01$.

Table A12: Tax Simulation Details

Tax	Marital Status (1)	Number of Dependents (2)	Spousal Income (3)	Control Mean (4)	LATE (5)	Estimated Share (6)
State + Federal + 19% FICA	Single	0	N/A	654	398	0.361
State + Federal + 19% FICA	Single	1	N/A	77	112	0.022
State + Federal + 19% FICA	Single	2	N/A	-290	-86	0.026
State + Federal + 19% FICA	Single	3	N/A	-510	-199	0.013
State + Federal + 19% FICA	Married	0	None	423	259	0.103
State + Federal + 19% FICA	Married	0	$15,000	805	434	0.103
State + Federal + 19% FICA	Married	0	Same	1,327	793	0.103
State + Federal + 19% FICA	Married	1	None	-81	18	0.032
State + Federal + 19% FICA	Married	1	$15,000	322	294	0.032
State + Federal + 19% FICA	Married	1	Same	930	662	0.032
State + Federal + 19% FICA	Married	2	None	-465	-171	0.039
State + Federal + 19% FICA	Married	2	$15,000	-125	136	0.039
State + Federal + 19% FICA	Married	2	Same	586	520	0.039
State + Federal + 19% FICA	Married	3	None	-679	-276	0.019
State + Federal + 19% FICA	Married	3	$15,000	-408	5	0.019
State + Federal + 19% FICA	Married	3	Same	344	401	0.019

Note: Column (4) reports estimated tax liabilities simulated from NBER TAXSIM 27 under the family structure and spousal income assumptions in columns (1) through (3). Estimates in column (5) report the impact of program enrollment on annual tax liabilities calculated from reported earnings under each set of family structure and spousal income assumptions. LATE estimates are calculated using equation (5) with estimated tax liabilities as the dependent variable in a longitudinal dataset that is unique at the individual-by-year level, with heteroskedasticity-robust standard errors clustered at the individual level. Proportions in column (6) are authors' calculations from ACS data describing the population of Framingham, MA (using 2017 ACS tables B05009, B09005, and S0501), assuming that spousal income is evenly split between the three categories for individuals who are married. Data restricted to 2010 to 2018 observations, the only years where full annual earnings are available (earnings data is only observed through quarter 3 of 2019). Spousal income categories of "None", "$15,000", and "Same" calculate household tax liabilities under the assumption that married couples file jointly and that household taxable earnings are equal to individual earnings ("None"), individual earnings plus $15,000 ("$15,000"), or twice individual earnings ("Same"). Alternative specifications that censor "Same" spousal earnings at $50,000 produce qualitatively similar results. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. N=20,059 annual earnings observations.

Table A13: Heterogeneity of First-Stage Estimates

	Male	Female	Beginner	Intermediate or Advanced	Brazilian Surname	Non-Brazilian Surname	Pre-Period Earnings > \$0	Pre-Period Earnings = \$0
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. First-Stage Measures								
Ever Enrolled at FAESL+	0.52	0.49	0.50	0.51	0.51	0.50	0.47	0.52
	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)
	1,929	2,832	4,191	570	2,093	2,688	531	2,643
Number of Terms Enrolled	1.8	1.5	1.6	1.8	1.6	1.6	1.6	1.5
	(0.2)	(0.1)	(0.1)	(0.2)	(0.2)	(0.1)	(0.2)	(0.1)
	1,929	2,832	4,191	570	2,093	2,688	531	2,643
Total Hours Enrolled	137	117	124	135	125	127	119	123
	(13)	(10)	(9)	(19)	(13)	(10)	(18)	(10)
	1,929	2,832	4,191	570	2,093	2,688	531	2,643

Note: Results are estimated using equation (2) in a dataset that is unique at the individual-level, with heteroskedasticity-robust standard errors in parentheses followed by the number of observations that contribute to each estimate. All outcomes defined over post-lottery periods only. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; and an indicator for missing gender. Beginner and Intermediate/Advanced subgroups are identified based on initial (entry) level of English. All p-values < 0.001.

Table A14: Effects on Annual Employer-Reported Earnings by Lottery Semester

	Control Mean (Annual Earnings through Y_{10}) (1)	Effect on Annual Reported Earnings (in \$) through Y_{10} (2)	Effect on Annual Reported Earnings (in \$) Y_2—Y_{10} (3)
Fall 2008	$2,692	2,256	2,256
		(2,172)	(2,172)
		3,672	3,672
Spring 2009	$3,808	3,996	4,312
		(2,967)	(3,095)
		3,699	3,288
Fall 2009	$3,706	3,723	3,820
		(4,180)	(4,322)
		3,105	2,760
Spring 2010	$4,187	-2,271	-2,278
		(1,854)	(1,995)
		3,807	2,961
Fall 2010	$3,909	343	-366
		(2,450)	(2,915)
		2,790	2,170
Spring 2011	$5,079	-914	-1,747
		(3,034)	(3,697)
		1,832	1,374
Fall 2011	$4,087	6,952	8,256
		(4,163)	(4,662)
		2,176	1,632
Spring 2012	$5,208	1,206	2,196
		(3,505)	(4,344)
		1,750	1,250
Fall 2012	$8,408	3,395	4,763
		(2,966)	(3,857)
		1,253	895
Spring 2013	$2,886	-633	-747
		(1,608)	(1,951)
		1,314	876
Fall 2013	$4,331	4,509	6,836
		(3,411)	(4,617)
		1,410	940
Spring 2014	$4,459	925	1,969
		(3,547)	(4,591)
		945	567
Fall 2014	$3,687	910	2204
		(2,472)	(3,517)
		1,345	807
Spring 2015	$2,580	2,701	4,272
		(2,070)	(2,665)
		916	458
Fall 2015	$4,388	1,049	1,676
		(1,372)	(2,048)
		1,508	754
Spring 2016	$3,112	-20	194
		(744)	(972)
		1,248	416

Note: Column (1) reports mean annual reported earnings through Year 10 among individuals in the indicated lottery cohort who did not win their first lottery attempt. Columns (2) and (3) report 2SLS IV estimates using equation (5) of the effect of enrolling in the FAESL+ program on reported earnings through Year 10 (Column 2) and from Year 2 through Year 10 (Column 3) for the lottery cohort, with heteroskedasticity-robust standard errors clustered at the individual-level in parentheses. All estimates include period fixed effects and the following covariates: Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; baseline quarterly earnings and an indicator for missing gender. All p-values are > 0.05.

Table A15: Effects on Civic Outcomes by Lottery Semester

	Control Mean (% Registered) (1)	Ever Registered to Vote (2)	Ever Voted (3)
Fall 2008	0.058	0.052	0.088
		(0.068)	(0.060)
		408	408
Spring 2009	0.070	0.053	0.140
		(0.077)	(0.079)
		411	411
Fall 2009	0.095	0.211	0.198
		(0.154)	(0.141)
		345	345
Spring 2010	0.091	0.081	-0.013
		(0.074)	(0.066)
		423	423
Fall 2010	0.046	0.120	0.109
		(0.099)	(0.087)
		310	310
Spring 2011	0.078	0.258*	0.115
		(0.107)	(0.080)
		229	229
Fall 2011	0.093	-0.069	-0.026
		(0.098)	(0.098)
		272	272
Spring 2012	0.090	0.153	-0.054
		(0.110)	(0.095)
		250	250
Fall 2012	0.071	0.066	0.151
		(0.109)	(0.111)
		179	179
Spring 2013	0.047	0.042	-0.005
		(0.073)	(0.082)
		219	219
Fall 2013	0.042	0.195	0.152
		(0.108)	(0.124)
		235	235
Spring 2014	0.047	0.267*	0.169
		(0.106)	(0.112)
		189	189
Fall 2014	0.041	0.079	0.086
		(0.077)	(0.077)
		269	269
Spring 2015	0.033	0.005	0.054
		(0.051)	(0.064)
		229	229
Fall 2015	0.033	-0.044	0.055
		(0.046)	(0.051)
		377	377
Spring 2016	0.029	0.007	0.040
		(0.042)	(0.053)
		416	416

Note: Column (1) reports the proportion of the control group who ever registered to vote after their first lottery application and before 2017. Columns (2) and (3) report 2SLS IV estimates using equation (3) of the effect of enrolling in the FAESL+ program on voter registration and voting. All estimates include covariates and lottery fixed effects that interact incoming level with time-of-day preferences and semester of first lottery application. Covariates include gender; Asian, Hispanic or white surname; Brazilian surname; surname not attributed to any racial or ethnic group; age at lottery; and an indicator for missing gender.