Do Language Requirements Deter Naturalization? Do They Promote Language Acquisition?

Jacob L. Vigdor

Duke University and NBER

INCOMPLETE Preliminary Draft

February 2010

Please do not cite or quote; comments welcome

Abstract

Since 1906, the United States has required most immigrants to pass an English language requirement to become citizens. The two most noteworthy exceptions apply to immigrants above a specific age who have spent a specific number of years in residence. This paper studies both the impact of the 1906 law, and the modern effect of exemptions, on the propensity of immigrants to learn English and naturalize. Results indicate that the language requirement has had a significant positive effect on English language ability, as well as a significant negative impact on naturalization rates.

Introduction

Since 1906, the United States has required applicants for naturalization to take a required citizenship examination in English. Later in the 20th century, Congress established two exemptions to this law. Foreign-born residents who have held permanent resident status in the United States for at least 20 years and reached the age of 50, as well as those at least 55 who have held permanent resident status for 15 years, are eligible to take the civics exam in a language of their choice. Other exemptions exist in cases of physical or mental disability.

These exemptions are controversial. There have been reports of fraud and abuse.[1] Commentators have assailed the exemptions as a symptom of "liberal victimology."[2] Proposals to eliminate the exemptions are often discussed in conjunction with more general efforts to elevate the official status of the English language. Before thinking seriously about any such proposals, however, it seems reasonable to ask a basic question: are these exemptions relevant? Is there a numerically significant population of foreign-born permanent residents who wait until they reach specific age or duration-of-residency thresholds before applying for citizenship, because they have no intention of learning English? Or is the language acquisition process robust enough to ensure that any interested resident probably knows enough English after 15 or 20 years to pass the exam anyway?

---

[1] See, for example, Michael Matza (August 27, 2009) "Exemption on Citizenship Test Can Be Abused." *Philadelphia Inquirer*.
[2] Ian de Silva (May 28, 2008) "Language Scofflaws: Congress' English Exemption." *Washington Times*.

This paper uses basic quasi-experimental techniques to infer the impact of the English language restrictions on transitions to citizenship. Two specific aspects of naturalization policy are used for purposes of inference. First, the introduction of the English language requirement in 1906, coupled with the standard 5-year waiting period for citizenship that applied to all immigrants at that time, implies that those immigrants who arrived after 1901 faced a higher barrier to citizenship than those who arrived in that year or earlier. Moreover, the barrier was erected several years after the immigration decision, which bolsters the argument that it was an exogenous policy change from the perspective of the individual immigrant.

The second source of identifying variation comes from the eligibility criteria for exemption from the English language requirements under current law. The exemptions, if valuable, should create discontinuities in naturalization rates at specific combinations of age and years in the United States, particularly among immigrants who report poor English skills.

Results indicate that the English language requirement indeed discourages some non-Anglophone immigrants from becoming citizens. The magnitude of the estimated effects are substantial in some cases. Estimates indicate that an immigrant born in a non-Anglophone country with a baseline likelihood of naturalization equal to 10% are half again as likely to become a citizen once they qualify for the age 55/15 years of residence exemption. Estimates do vary across specifications, and the 55/15 exemption is fairly consistently estimated to be more important than the 50/20 exemption. Estimates also indicate that the imposition of the English language requirement in 1906 delayed the acquisition of citizenship for many immigrants: the impact on citizenship in 1920 is

comparable to the estimated importance of the 55/15 exemption just reported; the magnitude of the effect is lower by 1930, when immigrants have had an extra ten years to learn sufficient English.

Further analysis shows that immigrants eligible to take the citizenship examination in their native tongue are significantly less likely to speak English in 1920, and self-report poorer English skills in 2007. These findings support the notion that the English language requirement acts as an incentive to acquire language skills.

Why language and naturalization matter

Immigrants' decisions to learn the native language of their host country, and to become citizens of that nation, can be modeled as investment decisions. In both cases, the immigrant undertakes actions with immediate costs that promise benefits in the future. As such, the standard economic models of human capital apply quite readily to both decisions. Generally speaking, under conditions of complete information and fully functioning markets, we would expect immigrants to optimally invest in both forms of human capital.

Public policy enters into the discussion of each of these decisions, for very different reasons. From an economic perspective, straightforward behavioral models introduce the possibility of positive externalities in language acquisition (Lazear 1995). The basic assumption generating this result is that the degree of surplus in market interactions between two agents is greater when the two agents speak the same language. Language is used to convey subtle signals about an agent's willingness to pay or accept, or about the quality of certain goods or services. When these signals cannot be conveyed, some otherwise beneficial transactions might not take place. So long as each party to the lost

transaction stood to gain some of this benefit, neither has appropriately strong incentives to acquire the language skills necessary to interact with the other. Recent literature in development and public economics has associated a host of negative societal outcomes with linguistic diversity (cites).

As with any positive externality, appropriate governmental response takes the form of subsidizing the externality-generating activity. Generally speaking, the government should subsidize the learning of additional languages. In theory, one could encourage the acquisition of a single common language by all residents of a society, or encourage the acquisition of multiple languages. In practice, the most cost effective means of promoting linguistic harmony would generally involve encouraging members of linguistic minority groups to acquire the dominant majority language.

Public policy plays a much more obvious role in the case of naturalization. Public policy establishes the costs of naturalization, in the form of waiting times, application fees, and tasks required to successfully complete the process. Policy also establishes the benefits, by enumerating the rights of naturalized citizens relative to those of non-citizens. Naturalization policy can itself be a policy lever; by offering the incentive of additional rights and privileges, governments can encourage immigrants to undertake acts that they might not otherwise consider.

Naturalization, then, can serve as the extra inducement necessary to correct the positive externality in language acquisition, under certain conditions. The ratio of benefits to costs associated with naturalization must be sufficiently high that the additional burden of learning English will not cause a wholesale abandonment of the process. Put differently, the cost of learning English must not be high relative to the net benefit associated with

naturalization. Economically speaking, the net benefits of naturalization can be used as a means of equating the private and social marginal benefits of language acquisition.

The main complication to this strategy is that the net benefits of naturalization vary from person to person. As the benefits accrue over time, the net present value depends on expectations regarding duration to retirement. The benefits also vary according to the differential in earnings expected in origin and host countries, and with the valuation of non-monetary amenities associated with citizenship. Altogether, then, it is reasonable to think that the offer of citizenship can be used as an incentive for certain types of behavior, but the actual construction of a Pigovian naturalization policy that exactly offset positive externalities in language acquisition is purely hypothetical.

Naturalization policy in the United States

Since the American colonial period, the naturalization process has evolved from an ad hoc matter undertaken by legislatures on behalf of one individual or family, to a multi-stage official process governed by numerous rules and regulations. In some sense, however, the modern naturalization process closely resembles the procedure set forth by the first Naturalization Act in 1790. For 220 years, residents wishing to naturalize have faced waiting periods, a "moral character" clause, and a requirement to renounce allegiance to foreign governments (which is in practice only sporadically enforced: dual citizenship is common among non-natives in the United States).

The Basic Naturalization Act of 1906 introduced the requirement that applicants for naturalized citizenship be able to communicate in English. Coupled with the standard five-year waiting period in place at the time, the Act implied that immigrants arriving in the

United States after 1901 would have no opportunity to become citizens without knowing English. Until 1952, naturalization law also incorporated a declaration of intent requirement. As of 1906, the waiting period following declaration of intent was two years. Thus non-English speaking immigrants arriving in 1901 or before who had not yet declared intentions by 1904 also lost the opportunity to become citizens without knowing English.

Basic evidence suggests that the imposition of the English language requirement had no appreciable effect on immigrants' propensity to naturalize. Figure 1 presents synthetic cohort evidence drawn from the US Census enumerations of 1900, 1910, 1920, and 1930. The solid line focuses on immigrants who reported arriving in the US between 1896 and 1900, indicating the proportion of this group who reported being naturalized citizens in each Census. Given the existence of the five-year waiting period, it is not surprising to note that the naturalization rate for this group was close to zero in 1900. Twenty years later, the rate was between 50 and 60 percent. Although these immigrants would have had to declare intentions to naturalize by 1904 in order to be exempt from the English requirement, every member of the cohort would have had an opportunity to qualify.

The dashed line shows the naturalization rate for immigrants reporting an arrival date between 1906 and 1910. Within this group, the English language requirement was clearly in effect. Nonetheless, the proportion of the group naturalizing by 1930 is 60% -- a rate slightly greater than that observed among their predecessors. There are many reasons to be skeptical of this simple comparison: the composition of the immigrant population may have changed across cohorts, and world conditions were very different in the 1920s

than they had been in the prior decade. Further analysis below will attempt to address these concerns.

The Immigration and Nationality Act (also known as the McCarran-Walter Act) of 1952 altered naturalization policy in many respects. For purposes of this analysis, the most important change was the establishment of an English language exemption, for immigrants over 50 with at least 20 years' residence in the United States. The second English exemption, for immigrants over the age of 55 with at least 15 years' residence, was enacted in 1990.

The McCarran-Walter Act abolished the declaration of intent requirement. The Act also, however, established legal permanent residency as a stepping stone to citizenship. Once an individual becomes a permanent resident, eligibility for citizenship comes after five years. The wait for legal permanent residency, among those immigrants eligible to receive it, can last a decade or more. This complicates efforts to connect time in the United States to propensity to become a citizen, as Census enumerations do not routinely collect information on the date that a foreign-born alien became a permanent legal resident. The uncertain length of the wait for a green card will serve as a significant caveat to analysis presented below.

Perhaps because of the English language restriction, the likelihood of naturalization among contemporary immigrants increases monotonically with self-reported English language skills. Figure 2 presents information gathered from the 2007 American Community Survey on immigrants above the age of 18. Only 18% of immigrants with no knowledge of English are naturalized citizens, compared to more than half of immigrants who speak English exclusively or "very well." This pattern could, of course, reflect factors

other than the English language requirement. Given that language ability tends to improve with time in the US, and the waiting periods built into the naturalization process, this pattern could be solely a function of time in the host country. Immigrants interested in becoming citizens might also naturally be more interested in learning English, whether it is required or not.

Data and methods

Inferring the impact of English language requirements on the propensity to naturalize, and the propensity to learn English itself, depends on the availability of certain data items and the existence of a causal identification strategy. Data on English ability and citizenship are critical to the exercise, and the ability to observe eligibility for exemptions will play a central role in the identification strategy.

The Census Bureau collected information on language ability and citizenship consistently between 1900 and 1930, and again from 1980 through 2000. Since 2000, the American Community Survey has continued to collect this information on an annual basis. In each of these years, the Bureau also collected information on the year that an immigrant arrived in the United States, as well as age information for all respondents. In the early 20th century, this information is sufficient to impute whether an immigrant would have had an opportunity to naturalize before the English language requirement went into effect in 1906. Coding of eligibility for exemptions in the later time period is imperfect, as we do not observe the year that an immigrant became a permanent resident. At least some immigrants receive green cards upon arrival in the United States, however, so the use of year of arrival as a proxy for year of permanent residence can be thought of as introducing

measurement error. As such, the results here will tend to understate the effect of the exemptions on naturalization.

The impact of both the 1906 law change and the later exemptions could be inferred with a simple regression-discontinuity style analysis, comparing the naturalization rates (and English-speaking ability) of immigrants barely eligible to obtain citizenship without learning English to those barely ineligible. To further test the robustness of findings, a falsification test will be added to most specifications. Certain immigrants should be entirely unaffected by the English language requirement – those who can speak English. Were language ability not potentially a function of naturalization policy itself, one could stratify the immigrant population by their ability to speak English. This specification will be used in some cases below.

A more defensible strategy, and one used in preferred specifications below, stratifies the immigrant sample by a more exogenous characteristic: birthplace. For immigrants born in Anglophone nations, the English language requirement is quite likely to be irrelevant. This is not always true; there are some nations, such as Canada, where English is predominant but not the exclusive native tongue. Birth in an Anglophone nation should be highly correlated with irrelevance of the English language requirement, however. The basic empirical strategy below, then, will be to evaluate the citizenship rates and English language abilities of immigrants based on their eligibility to become citizens without learning English. The identification strategy takes advantage of the sharp cutoff rules for exemptions, as well as the hypothesized differential impact on immigrants from Anglophone nations (or in some cases, for immigrants with higher-quality English skills). A representative estimating equation takes the following form:

(1) $y_{ijt} = f(age_{ijt}) + g(years\ since\ immigration_{ijt}) + b_1 * e_{ijt} + b_2 * a_j + b_3 * e_{ijt} * a_j + u_{ijt}$,

where y_{ijt} is the outcome of interest for immigrant i born in nation j in year t – a measure either of citizenship or of English ability, $f()$ and $g()$ are smooth polynomial functions of the indicated variables, e_{ijt} indicates eligibility for an exemption from the English requirement, a_j indicates whether j is an Anglophone nation, and u_{ijt} is an error term. Although written as a linear equation, in reality estimation below will make use of the probit functional form. In this scenario, y_{ijt} can be thought of as a latent indicator that controls the value of the binary (or ordinal) outcome of interest, and u_{ijt} is presumed to follow a standard normal distribution.

Naturalization is a consistently measured binary outcome. The form of data collection used by the Census Bureau changed between 1900 and 2007, switching from an in-person interview to a mail questionnaire. Throughout the time period, though, immigrants have been asked to self-report their citizenship status. The Census Bureau has never had a policy of verifying this information, though it is conceivable that some enumerators would ask for evidence. As such, this entire analysis is subject to the caveat that respondents may have systematically misreported their citizenship status, perhaps because they feared negative consequences of revealing their true status.

The measurement of English language ability has changed more substantially over time. Between 1900 and 1930, enumerators had the ability to verify whether a respondent spoke English, and coded this information as a simple binary indicator. Since 1980, English language ability has been self-reported, and respondents can choose from a number of ordinal categories, from "does not speak English" to "speaks only English." As such, analysis of language ability in 2007 will be conducted using an ordered probit model. The

switch to self-reported data introduces concerns about misreporting, which cannot be fully addressed in this analysis.

Results

Effects of the English language exemptions

At heart, the identification strategy used to study the modern-day exemptions to the English language requirement is a basic difference-in-difference-in-difference model: comparing immigrants of varying ages, with different durations of stay in the United States, and different English ability. Tables 1a and 1b present the most basic evidence using this strategy, simple mean differences comparing immigrants in different categories.

Table 1a focuses on the age 50/20 years' permanent residence exemption. The first entry in the table shows the baseline naturalization rate for immigrants who are under 50 and have been in the United States for less than 20 years. It then examines the difference in naturalization rates among immigrants who qualify along one dimension but not the other. From a base naturalization rate of 23%, the boost to naturalization from attaining 50 years of age alone is 14 percentage points, and the boost associated with reaching 20 years' duration in the United States is a more substantial 41.6%. Adding these three numbers provides a simple prediction of the naturalization rate among immigrants who qualify for both exemptions. This prediction appears next in the table, followed by the actual observed rate among immigrants who qualify for both exemptions. Were the actual figure to exceed the additive prediction, we might attribute the difference to the effect of qualifying for the exemption. When studying the entire immigrant population, there is

little difference between the predicted and actual rates, suggesting that the overall impact of the exemption in the entire immigrant population is minimal.

The remaining rows of the table break the immigrant population down according to self-reported English ability. When analyzed in this manner, a different picture appears. Among those immigrants who self-report speaking no English, the actual naturalization rate among immigrants who qualify along both dimensions is 17.2 percentage points higher than would be predicted on the basis of an additive prediction of qualifying along one dimension only. By contrast, the prediction is fairly close to the actual rate for other groups, those who speak English at least "not well." The difference-in-difference-in-difference estimator, comparing non-speakers to those who speak English "very well," is 19.2 percentage points – the naturalization rate among non-English speaking immigrants who qualify for the exemption is roughly two-thirds higher than we would otherwise expect.

A similar pattern appears in Table 1b. In the immigrant population overall, there appears to be little impact of qualifying for the age 55/15 years permanent residence exemption. When broken down by self-reported English ability, however, there is once again a sizable difference between the naturalization rate observed for non-speakers who qualify for the exemption and what would be predicted based on an addition for qualifying along each dimension in isolation. The difference-in-difference-in-difference estimate in this case is 15.8 percentage points.

The difference-in-difference-in-difference methodology can be misleading if the effects of age and duration of stay in the United States have a nonlinear impact on the propensity to become a naturalized citizen. The regression discontinuity framework

addresses this concern to a large extent, by basing identification of effects on the comparison of respondents barely eligible or ineligible for an exemption. As noted above, full eligibility is impossible to determine, as the American Community Survey contains no information about the date at which an immigrant became a permanent resident. This analysis must therefore be considered a "fuzzy" regression discontinuity. In such cases, the reduced-form treatment effect obtained from a comparison of individuals on either side of the eligibility cutoff must be scaled up according to the difference in actual eligibility. Actual eligibility is not observed in these data; thus the most accurate possible statement is that any effects uncovered here understate, by an unknown amount, the true causal effect of eligibility.

Table 2 presents the results of basic parametric RD specifications, which control for quartic polynomials in immigrants' age and years since immigration. Each entry in this table represents an RD estimate taken from a separate specification. Across rows, the estimates differ by the sample used to estimate them. The left column shows estimated effects of the age 50/20 years' residence exemption, while the right column shows estimated effects of the age 55/15 years' residence exemption. All specifications are estimated using a probit regression.

When estimated on the full sample of adult foreign-born residents, the estimated effects of eligibility for either exemption are, surprisingly, negative – indicating that those individuals barely eligible for the exemption are actually less likely to be naturalized than those barely ineligible. The estimated effect is sizable in the case of the 50/20 exemption: relative to a barely-ineligible respondent with a 50% likelihood of being naturalized,

equivalent barely-eligible respondents naturalize at only a 45% rate. The estimated effect of the 55/15 exemption is more modest.

These surprising negative effects are also present when the sample is restricted to immigrants from non-Anglophone nations, though only the 50/20 estimate is statistically significant. Stratifying the sample by self-reported English ability, we observe negative point estimates for those in the lowest two categories in the case of the 50/20 exemption. For immigrants who either speak no English or report speaking English "not well," there are positive and significant point estimates for the effect of eligibility for the 55/15 exemption. Compared to a barely-ineligible non-English speaking immigrant with a baseline 25% likelihood of naturalization, immigrants barely eligible for the 55/15 exemption have a 29% naturalization rate.

While the RD specifications in Table 2 better address concerns about nonlinear age and duration effects, they do not mimic Table 1's triple-difference method. Table 3 presents probit estimates derived from the entire sample of foreign born adults in 2007, which introduce the third dimension of difference: comparing the impacts of eligibility for immigrants of varying English-speaking status. The first and third specifications control for the variables depicted as well as quartic polynomials in age and duration of stay in the United States. The second and fourth specifications add a supplemental set of covariates, including indicator variables for country of birth, educational attainment, gender, and race.

The first rows in Table 3 report main effects, which show several interesting patterns. Unsurprisingly, the propensity to naturalize among immigrants not eligible for any exemption increases substantially with self-reported English ability; immigrants from Anglophone nations are also significantly more likely to become citizens than otherwise

similar individuals from non-Anglophone nations. More surprisingly, eligibility for both the 50/20 and 55/15 exemptions are associated with significantly lower rates of naturalization among immigrants who speak English exclusively, or among immigrants born in Anglophone nations. There is no obvious explanation for this pattern; the quartics in age and duration of stay should control flexibly for smooth trends in those variables.

Setting aside these puzzling main effects for a moment, the remainder of the table shows a fairly consistent pattern of significant interaction effects: immigrants born in non-Anglophone countries, or who self-report poor English speaking ability, become significantly more likely to naturalize when they become eligible for an exemption to the English language requirement. In the most basic specifications in columns 1 and 3, the strongest effects are associated with the 55/15 exemption. The estimates are of substantial magnitude. The triple-difference estimates suggest that relative to a barely-ineligible immigrants with a 25% likelihood of being naturalized, one just barely eligible for the 55/15 exemption has a 46% likelihood. Factoring out the significant main effect, the estimated change in naturalization rates is more modest, for example from 25% to 33%. Estimates shrink somewhat when additional covariates are added to the specification. Nonetheless, eligibility for exemptions continues to predict a noticeable increase in the likelihood of naturalization for immigrants whose native language is not English, relative to native speakers. Notably, in the specification examining differential impacts on immigrants from non-Anglophone countries, the 50/20 exemption becomes a more significant predictor of naturalization. The pattern of significant negative main effects – declines in naturalization rates among native speakers upon eligibility for exemptions – persists in these specifications.

Effects of the 1906 English requirement

The 1906 English language requirement, coupled with the five-year waiting period for citizenship in effect at the time, created additional barriers to naturalization for non-English speaking immigrants who arrived after 1901. Table 4 shows simple difference-in-difference estimates, comparing the naturalization rates of pre- and post-1901 arrivals, distinguished by whether they were born in an Anglophone country. The estimates are presented for two points in time: 1920 and 1930. Naturalization rates are compute using only those immigrants who report having arrived in the United States at least five years before enumeration.

In all cases, more recently-arrived immigrants are less likely to be naturalized citizens. The difference in rates before and after 1901 is more dramatic for immigrants born in non-Anglophone countries. In 1920, the difference-in-difference estimates indicates that the English language requirement reduced naturalization rates for immigrants in this group by 10.4 percentage points. By 1930, rates for recently-arrived immigrants from Anglophone and non-Anglophone nations had converged somewhat, leaving only a 3.6 percentage point difference-in-difference. There are two straightforward explanations for the decline in the law's impact. First, immigrants may have learned English over time, making the restriction irrelevant. Second, Anglophone immigrants undoubtedly arrived in greater numbers during the 1920s, as country-of-origin based quotas biased against immigrants from Southern and Eastern Europe took effect. These hypotheses are distinguishable in the more detailed analysis below.

Table 5 shows a series of regression discontinuity estimates to evaluate the impact of the 1906 English language requirement. As in Table 2 above, each table entry comes from a separate probit regression which controls for a quartic in years since arrival in the US. Estimates based on 1920 data show a consistent negative discontinuity, implying that immigrants arriving after 1901 had lower naturalization rates. The effect is concentrated among immigrants who report not speaking English, or who were born in non-Anglophone nations. By 1930, however, the point estimates have either become significantly positive (for the overall population or immigrants born in non-Anglophone nations) or reverted towards zero (for non-English speakers). The significant positive effects appear to be concentrated in the segment of the population born in Anglophone nations, which hints that a difference-in-difference style estimator will continue to show a negative impact of arrival after 1901 for non-native English speaking immigrants.

Table 6 shows that this inference is in fact correct, in specifications that control for a quartic in years since arrival as well as additional categorical controls for country of birth, sex, and race. Interestingly, the difference-in-difference estimates are significant in the expected direction only when stratifying immigrants by Anglophone country of birth. When stratifying by English ability, point estimates are wrong-signed in both cases and significant in the 1920 specification.

Does naturalization policy incentivize language acquisition?

In theory, the benefits associated with naturalization, coupled with the English language requirements, provide most immigrants with an incentive to learn English. Table 7 analyzes the strength of these incentives, utilizing variation in the strength of incentives

associated both with the imposition of language requirements in 1906 and with exemptions in place in 2007.

The first set of specifications is estimated using a probit specification and data on immigrants born in non-Anglophone nations, from the 1920 Census. Controlling for a quartic in years since arrival, point estimates indicate that immigrants who arrived in 1901 were significantly less likely to be coded as English speakers in 1920, compared to those who arrived just afterward. Adding additional covariates to the analysis, the magnitude of the effect declines somewhat, but retains statistical significance. Relative to an immigrant arriving in 1901 with a 50% chance of being an English speaker in 1920, a comparable immigrant arriving in 1902 had a 52.2% chance of being an English speaker.

The third and fourth columns in Table 7 analyze immigrants' self-reported English ability in 2007, using an ordered probit. In both specifications, which vary by inclusion of additional characteristics beyond quartics in age and years since immigration, immigrants who qualify for either the 50/20 or 55/15 exemption report having significantly poorer English language skills. In the more complete specification, both exemptions have a roughly similar magnitude. Altogether this evidence, spanning more than a century, indicates that the English language requirement encourages some immigrants to improve their English skills.

Conclusions

This paper presented evidence on the importance of the English language requirement embedded in American naturalization policy since 1906. Results indicate that the requirement is a binding constraint for many immigrants, and encourages others to

improve their language skills.

Table 1a: Basic diff-in-diff-in-diff estimates of the impact of the 50/20 exemption

Group	Baseline	+over 50	+20 years in US	Additive prediction	Actual rate	Difference
All immigrants	23.0%	+14%	+41.6%	78.6%	77.0%	-1.6%
Self-reported English speaking:						
None	3.3%	+14%	+12.5%	29.8%	47.0%	+17.2%
"not well"	10.6%	+27.8%	+27.9%	66.3%	66.5%	+0.2%
"well"	26.7%	+25.1%	+35.4%	87.2%	82.2%	-5.0%
"very well"	35.9%	+14.8%	+37.7%	88.4%	86.4%	-2.0%
DDD: non-English speakers vs. those who speak "very well"						+19.2%

Source: 2007 American Community Survey, author's calculations. Proportions are estimated using ACS sampling weights.

Table 1b: Basic diff-in-diff-in-diff estimates of the impact of the 55/15 exemption

Group	Baseline	+over 55	+15 years in US	Additive prediction	Actual rate	Difference
All immigrants	17.2%	+11.3%	+41.6%	70.1%	77.0%	+6.9%
Self-reported English speaking:						
None	2.8%	+11.6%	+11.3%	25.7%	48.2%	+22.5%
"not well"	8.8%	+24.5%	+25.6%	58.9%	69.8%	+10.9%
"well"	21.5%	+20.6%	+36.9%	79.0%	83.9%	+4.9%
"very well"	27.5%	+9.9%	+42.5%	79.9%	86.6%	+6.7%
DDD: non-English speakers vs. those who speak "very well"						+15.8%

Source: 2007 American Community Survey, author's calculations. Proportions are estimated using ACS sampling weights.

Table 2: Basic Regression Discontinuity-style estimates of exemption effects

Sample selection criteria	Effect of 50/20 exemption	Effect of 55/15 exemption
All foreign-born over 18 (n=293,977)	-0.120**	-0.034*
	(0.013)	(0.014)
Foreign-born from non-Anglophone nations (n=264,341)	-0.139**	-0.020
	(0.014)	(0.015)
Foreign-born who self-report no English (n=30,129)	-0.036	0.134**
	(0.047)	(0.043)
Foreign-born who self-report speaking English "not well" (n=56,102)	-0.103**	0.071*
	(0.030)	(0.030)
Foreign-born who self-report speaking English "well" (n=64,006)	0.019	0.054
	(0.030)	(0.033)
Foreign-born who self-report speaking English "very well" (n=93,343)	0.035	0.001
	(0.027)	(0.031)

Note: Sample derived from the 2007 ACS. Standard errors in parentheses. Each row presents coefficients from a probit regression controlling for quartics in age and years since arrival in the US. Regressions are weighted using ACS sampling weights.
** denotes a coefficient significant at the 1% level; * the 5% level.

Table 3: Probit estimates using the full sample, 2007

Independent variable	Dependent variable: respondent is a citizen			
MAIN EFFECTS				
English ability (non-speaker omitted)				
Speaks English exclusively	1.28**	0.891**	---	---
	(0.020)	(0.025)		
Speaks "very well"	1.37**	0.872**	---	---
	(0.019)	(0.022)		
Speaks "well"	1.11**	0.752**	---	---
	(0.019)	(0.022)		
Speaks "not well"	0.545**	0.379**	---	---
	(0.020)	(0.022)		
Origin country is non-Anglophone	---	---	-0.102**	---
			(0.014)	
Exemptions (not qualified omitted)				
Qualifies for 50/20 exemption	-0.151**	-0.148**	-0.177**	-0.179**
	(0.028)	(0.029)	(0.033)	(0.035)
Qualifies for 55/15 exemption	-0.357**	-0.171**	-0.257**	-0.112**
	(0.029)	(0.030)	(0.034)	(0.035)
INTERACTION TERMS				
Non-speaker*50/20 eligible	-0.020	0.082	---	---
	(0.043)	(0.045)		
Non-speaker*55/15 eligible	0.541**	0.226**	---	---
	(0.043)	(0.045)		
Speaks very well*50/20	0.151**	0.138**	---	---
	(0.035)	(0.036)		
Speaks very well*55/15	0.127**	0.015	---	---
	(0.037)	(0.038)		
Speaks well*50/20	0.105**	0.147**	---	---
	(0.036)	(0.037)		
Speaks well*55/15	0.377**	0.161**	---	---
	(0.038)	(0.039)		
Speaks "not well"*50/20	-0.003	0.084*	---	---
	(0.036)	(0.037)		
Speaks "not well"*55/15	0.579**	0.281**	---	---
	(0.037)	(0.038)		
Non-anglophone*50/20	---	---	0.066	0.141**
			(0.034)	(0.036)
Non-anglophone*55/15	---	---	0.251**	0.086*
			(0.035)	(0.037)
Supplemental covariates	No	Yes	No	Yes
N	293,977	293,938	293,977	293,938

Note: Standard errors in parentheses. Table entries are probit coefficients. Data source is the 2007 ACS. Supplemental covariates include categorical controls for gender, race, educational attainment, and country of birth.

** denotes a coefficient significant at the 1% level; * the 5% level.

Table 4: Simple diff-in-diff estimates of the impact of the 1906 English requirement

Proportion of adults naturalized in 1920:	Arrived 1901 or earlier	Arrived after 1901	Difference
Anglophone origin country	83.5%	44.2%	39.3%
Non-anglophone origin country	74.0%	24.3%	49.7%
Diff-in-diff			10.4%
Proportion of adults naturalized in 1930:	Arrived 1901 or earlier	Arrived after 1901	Difference
Anglophone origin country	84.6%	55.3%	29.3%
Non-anglophone origin country	81.0%	48.1%	32.9%
Diff-in-diff			3.6%

Source: Public Use Microdata, US Census. Proportions estimated using IPUMS sampling weights. Sample excludes individuals arriving in the US within 5 years of the Census date, who are ineligible for naturalization.

Table 5: Basic RD-style estimates of the impact of the 1906 English requirement

Sample selection criteria	Discontinuity in naturalization rate at arrival date 1901
All foreign-born over 18, 1920 Census (n=131,650)	-0.044**
	(0.017)
All foreign-born over 18, 1930 Census (n=135,592)	0.083**
	(0.018)
Born in non-Anglophone countries, 1920 Census	-0.057**
(n=110,514)	(0.018)
Born in non-Anglophone countries, 1930 Census	0.056**
(n=107,191)	(0.020)
Non-English speakers, 1920 Census (n=16,735)	-0.213**
	(0.056)
Non-English speakers, 1930 Census (n=11,523)	0.018
	(0.070)

Note: Standard errors in parentheses. Each row presents coefficients from a probit regression controlling for quartics in age and years since arrival in the US. Regressions are weighted using IPUMS sampling weights.
** denotes a coefficient significant at the 1% level; * the 5% level.

Table 6: Probit estimates using the full sample, 1920 and 1930

Independent variable	1920 Census		1930 Census	
MAIN EFFECTS				
English ability (speaker omitted)				
Does not speak English	-0.501**	---	-0.948**	---
	(0.018)		(0.022)	
Exemptions (not qualified omitted)				
Arrived before 1901	0.004	-0.201**	0.149**	0.100**
	(0.017)	(0.024)	(0.019)	(0.025)
INTERACTION TERMS				
Non-speaker*arrived before 1901	-0.059*	---	-0.019	---
	(0.029)		(0.038)	
Non-anglophone*arrived before 1901	---	0.249**	---	0.075**
		(0.020)		(0.020)
Supplemental covariates	Yes	Yes	Yes	Yes
N	131,495	131,495	135,520	135,520

Note: Standard errors in parentheses. Table entries are probit coefficients. Data source is the 2007 ACS. Supplemental covariates include categorical controls for gender, race, and country of birth. Country of birth indicators subsume the main effect for non-anglophone origin.
** denotes a coefficient significant at the 1% level; * the 5% level.

Table 7: Does Naturalization Policy Incentivize Language Acquisition?

Independent variable	1920 Census Probit: speaks English		2007 ACS Ordered Probit: self-rated English ability	
Arrived before 1901 (had opportunity to attain citizenship before English requirement imposed in 1906)	-0.137** (0.025)	-0.054* (0.027)	---	---
Eligible for 50/20 exemption	---	---	-0.153** (0.012)	-0.159** (0.003)
Eligible for 55/15 exemption	---	---	-0.058** (0.013)	-0.151** (0.003)
Supplemental covariates	No	Yes	No	Yes
N	100,672	100,672	239,886	239,886

Note: Standard errors in parentheses. Table entries are probit coefficients. Data source is the 2007 ACS. All specifications include quartics in age and years in the United States. Supplemental covariates include categorical controls for gender, race, and country of birth. In 2007, supplemental covariates also include categorical controls for educational attainment. Country of birth indicators subsume the main effect for non-anglophone origin. Sample is restricted to adult immigrants born in non-Anglophone nations.
** denotes a coefficient significant at the 1% level; * the 5% level.

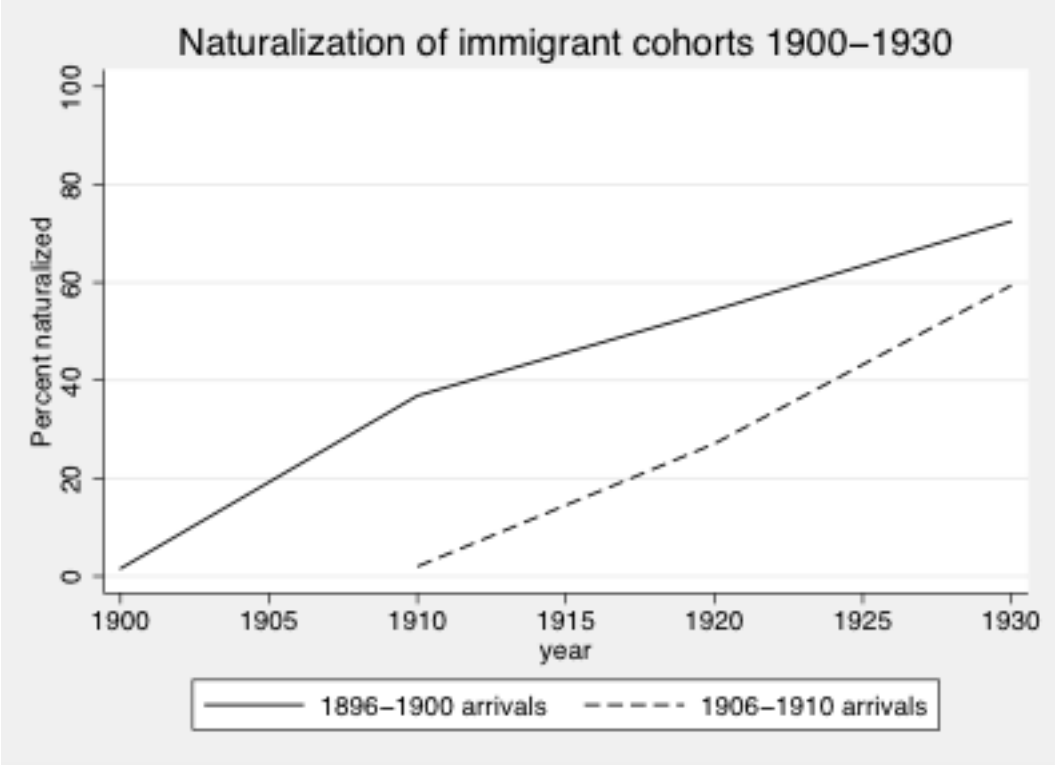
Naturalization of immigrant cohorts 1900–1930
Percent naturalized
0
20
40
60
80
100
1900
1905
1910
1915
1920
1925
1930
year
1896–1900 arrivals
1906–1910 arrivals

Figure 1

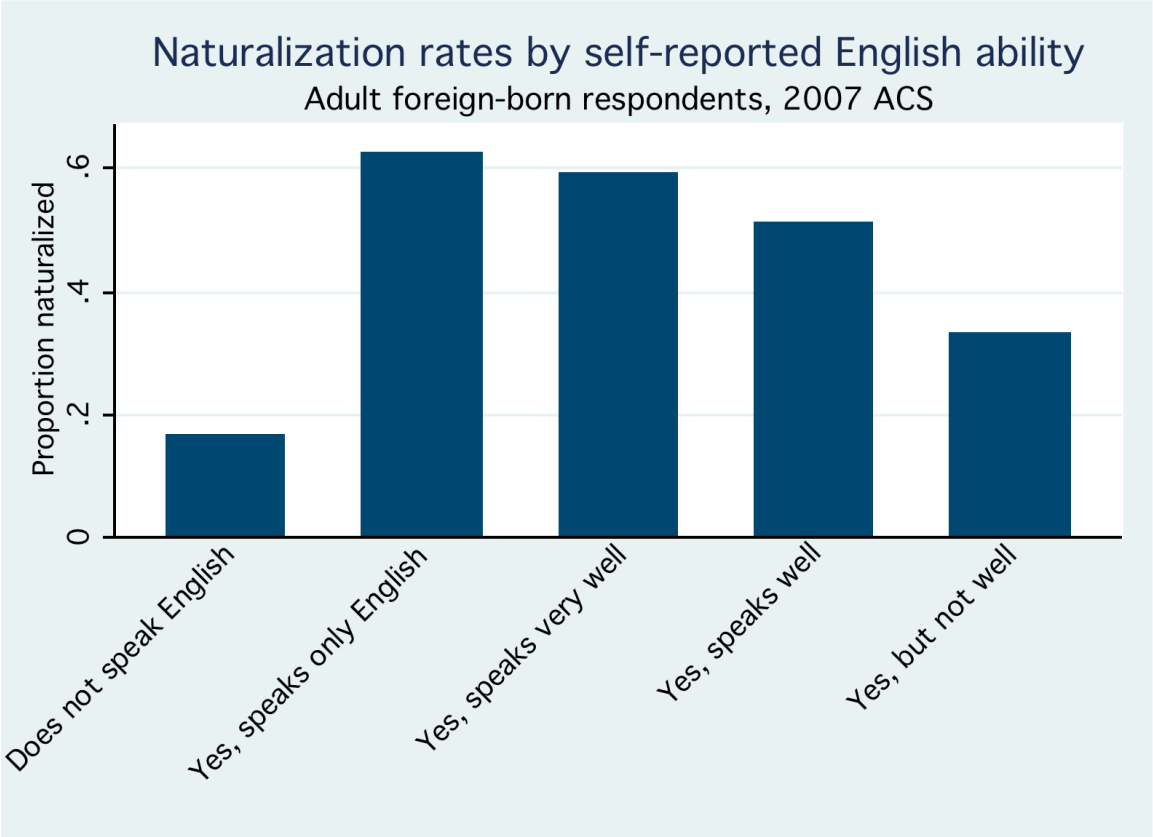
Naturalization rates by self-reported English ability
Adult foreign-born respondents, 2007 ACS
Proportion naturalized
0
.2
.4
.6
Does not speak English
Yes, speaks only English
Yes, speaks very well
Yes, speaks well
Yes, but not well

Figure 2