

IZA DP No. 9307

The Academic Progress of Hispanic Immigrants

Marie C. Hull

August 2015

The Academic Progress of Hispanic Immigrants

Marie C. Hull

*University of North Carolina at Greensboro
and IZA*

Discussion Paper No. 9307
August 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The Academic Progress of Hispanic Immigrants*

Past research has shown that Hispanic students make test score gains relative to whites as they age through school; however, this finding stands in contrast to the experience of blacks, who show little change in their relative position over the same time frame. Distinguishing Hispanic students by immigrant generation, I find that the children of immigrants (first- and second-generation Hispanics) drive the improvement in Hispanic test scores. Later-generation Hispanics consistently perform slightly below whites, perhaps due to negative selection into ethnic identification. Thus, previous estimates vastly understate the progress of first- and second-generation Hispanic immigrants. From a negative gap in 3rd grade, these students surpass socioeconomically similar whites in math and reading by middle school and end 8th grade as much as a quarter of a standard deviation ahead. Assimilation alone cannot explain this progress; a potential explanation is that immigrant parents create a home environment that fosters achievement.

JEL Classification: J24, I24, J15

Keywords: human capital, achievement gap, Hispanic immigrants

Corresponding author:

Marie Hull
Department of Economics
University of North Carolina at Greensboro
PO Box 26170
Greensboro, NC 27402-6170
USA
E-mail: mchull2@uncg.edu

* I would like to thank the members of my dissertation committee, Hugh Macartney, Arnaud Maurel, Seth Sanders, and especially my advisor, Peter Arcidiacono, for advice and encouragement. I am also grateful to the North Carolina Research Data Center and the North Carolina State Center for Health Statistics for access to the data.

1 Introduction

At 16.3% of the population, Hispanics are now the largest and fastest growing minority in the United States, yet they have lower levels of human capital than whites by several measures.¹ Furthermore, due to high fertility rates and high rates of migration into the country, Hispanics are disproportionately young, implying that a high proportion are currently receiving their education in our nation's schools. Given that they will be a key part of our future workforce, ensuring that Hispanic youth leave school with the skills to succeed has never been more important, yet Hispanic-white differences in human capital are almost always discussed as an afterthought to black-white differences in research on racial and ethnic inequality.

In recent work, Clotfelter, Ladd, and Vigdor (2009) document that of the three largest racial/ethnic groups (white, black, and Hispanic), Hispanic students are the only one that change their relative position in test scores over time. They score about 0.1 standard deviations below observably similar whites in 3rd grade, equal them in 5th, and outscore them by 0.1 standard deviations in 8th. In contrast, black students' covariate-adjusted reading gap stands at a constant -0.5 standard deviations; their math gap is similar in size, though it does narrow about a tenth of a standard deviation between 3rd and 8th grades. Since IQ is relatively stable by age 10, changes in underlying cognitive ability cannot explain the rise in Hispanic achievement over this time period, leaving noncognitive inputs and environmental factors as potential explanations (Cunha et al., 2006).

What, or rather, who, accounts for the relative gains in Hispanic test scores? Most papers that note an upward trend in Hispanic test scores attribute it to the assimilation of immigrant Hispanics over the sample period.² But in reality, only 13% of the school-age

¹According to the definition used in the 2010 Census, "Hispanic or Latino" refers to a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race. There is some variation in this definition by source and over time.

²In this paper, "assimilation" takes on two slightly different meanings. Here, and through most of the paper, I refer to the cultural process of assimilation, whereby immigrants adopt the native language and customs. Assimilation in test scores, like assimilation in wages, occurs when immigrant and native test scores are indistinguishable.

Hispanic population are immigrants, 42% have immigrant parents but were born in the U.S., and the remaining 45% come from families that have been in the U.S. for generations (Ruggles et al., 2010). Any measurement of human capital accumulation for Hispanics is a weighted average of these three groups. Moreover, any representation of how Hispanic students fare as they age blends together both the intercepts and slopes of each of these groups. In this light, lumping all Hispanics together or referring to all Hispanics as immigrants may be misleading. Although assimilation, whether it be through language acquisition or the adoption of native customs, might explain convergence in Hispanic test scores, it does not explain why Hispanic students outscore their native peers.

Several plausible scenarios could give rise to the upward trend in Hispanic test scores. I illustrate three of these in Figure 1, using the immigrant status of the parents to divide the Hispanic population into two groups. While each immigrant generation likely has its own intercept, they might share a common slope if Hispanic families are uniformly better at producing academic achievement. I show this scenario in Panel A. This case leaves no room for differential gains for immigrant Hispanics through the assimilation process.

It could also be the case that Hispanics from established families exhibit almost no change in test scores, like their white and black peers. This scenario would leave immigrant Hispanics responsible for all of the gains. Even here, the overall trendline could mask different levels of academic achievement for immigrant and non-immigrant Hispanics. In Panel B, I depict the case in which positively selected immigrant ancestors give later-generation Hispanics a lingering advantage in school. Then, the progress made by more recently arrived immigrants, though substantial, would not be enough to completely close their gap with whites. In Panel C, later-generation Hispanics appear slightly worse off compared to whites due to negative selection into identification as Hispanic, but more recent arrivals catch up quickly and go on to outscore whites. (I provide more detail on how each of these possibilities could come about in the following section.) In these latter two scenarios, the overall Hispanic trendline understates the gains made by immigrant Hispanic students. In the final one, their

test scores are also much higher than previously thought, especially as the students age.

Recognizing the important role of generational status, I decompose the Hispanic-white test score gap across grades. I find that the last scenario, depicted in Panel C of Figure 1, holds true. To estimate these gaps, I make use of a rich, administrative data set from North Carolina public schools that contains reading and math scores, as well as socioeconomic background information, for several cohorts of students as they progress from 3rd to 8th grade. Information on immigrant generation for students born in-state comes from matched birth records, and I use a finite mixture distribution to model test score production for unmatched students. Longitudinal data is critical in this type of study so as not to confuse changes in cohort quality with the speed of assimilation, a point first made by Borjas (1985).

To the best of my knowledge, no other paper has pinned down trends in test score gaps by immigrant generation with the level of precision found here. The contribution of this paper is threefold: First, I confirm that newly arrived Hispanic immigrants are responsible for the downward trend in Hispanic scores found with repeated cross sections. Previous work could not rule out the possibility that all Hispanic students are more sensitive to moves. Second, I find that later-generation Hispanics, like whites and blacks, consistently score slightly below socioeconomically similar whites. Although there is no evidence that their scores catch up to whites, this finding is of no real economic concern if the Trejo critique holds: An accurate measure of ethnic background would put them on par with whites. Most importantly, I show that the children of immigrants drive the growth in Hispanic test scores and that the speed of convergence for immigrant Hispanics is quite rapid. First-generation Hispanics that arrive by 3rd grade equal similar whites within a few years. The second generation makes similarly impressive gains to finish 8th grade 0.11 standard deviations ahead of whites in reading and 0.24 standard deviations ahead in math. Lumping all Hispanic students together understates the considerable progress of these immigrant students. While Borjas (1985) found assimilation in wages to be quite slow, my analysis shows that young immigrants, as well as the children of immigrants, catch up to native whites within a few years.

Two features of my results indicate that cultural assimilation is not the only force at work in the growth in Hispanic test scores. First, second-generation students make the same gains in math and close to the same gains in reading as the first generation. Though there is likely some room for the language skills of these second-generation students to catch up with whites, especially if Spanish is spoken in the home, they still have more years of U.S. experience overall. Second, the test scores of whites and immigrant Hispanics more than converge—second-generation Hispanics perform significantly better than whites in reading and math in 8th grade, the first generation for math only. If Hispanic immigrants adopting native customs was the sole factor behind the growth in Hispanic test scores, we would expect the growth to taper off as Hispanic scores approached average white scores.

Achievement test scores are determined by cognitive inputs, noncognitive inputs, and environmental factors. In terms of the rise in Hispanic immigrant achievement found here, Cunha et al. (2006) rule out changes in IQ after age 10, and controls for school quality rule out the school environment. Thus, noncognitive skills (e.g., motivation and self-discipline) and the home environment must drive the test score gains for children of immigrants. Although I do not have direct evidence, an explanation that is consistent with my findings is that immigrant parents create a home environment that values and fosters achievement. This environment gives their children a boost beyond any effects of assimilation. In any case, Hispanic immigrant students fare quite well under the current system, even better than previously suggested. While these students are at a severe socioeconomic disadvantage, their home environment and noncognitive abilities combine to lift them up above the achievement of similar whites.

In the following section, I discuss in greater detail the processes of immigration, assimilation, and selection that could generate the overall Hispanic trend. I also provide an overview of previous work on the Hispanic-white achievement gap and the role of generational status in educational outcomes for immigrants. In Section 3, I describe the data sets used in this paper, which motivate the development of my econometric framework in Section 4. I report

results in Section 5, detailing trends in raw and adjusted Hispanic-white test score gaps by immigrant generation as students move through school. Section 6 concludes.

2 Theoretical and empirical background

2.1 Immigration, assimilation, and selection

In their paper on minority achievement gaps, Clotfelter, Ladd, and Vigdor (2009) were the first to find that Hispanic students improve relative to whites with continuous enrollment over the course of grades 3 to 8. In fact, Hispanics score the same as whites in math and reading in 5th grade after adjusting for covariates, and they outperform them by about a tenth of a standard deviation in 8th grade. That Hispanics perform more poorly than whites of comparable socioeconomic status early in school but outscore them in later grades runs counter to the experience of black students. As shown by Clotfelter, Ladd, and Vigdor (2009), among others, black children also have lower test scores in the early grades, but there is no evidence of convergence as the students age.³ Furthermore, the enormous improvement in Hispanic achievement in these grades cannot be due to changes in cognitive ability, in accordance with the human development literature cited in the review article by Cunha et al. (2006), which states that there are critical periods of development and that IQ becomes stable by age 10.

How do we interpret the relative success of Hispanic children? One interpretation is that Cunha et al. (2006) are wrong and that the underlying cognitive ability of Hispanic children improves relative to their non-Hispanic peers well beyond age 10. Another interpretation, however, recognizes that Hispanics are a heterogeneous group, most importantly heterogeneous in how recently their families migrated to the U.S. Some Hispanic children are themselves immigrants, some are U.S.-born but have immigrant parents, and some are

³Fryer and Levitt (2004, 2006) find that the test scores of black Kindergarteners are no different from socioeconomically similar whites, but the difference opens up to commonly observed levels by 3rd grade.

children of Hispanic ancestry whose families have been in the U.S. for three or more generations. Thus, in every grade, the average test score of Hispanic students is a mixture of the performance of these three groups. Differences in the home environments and noncognitive skills among these groups of Hispanics could combine to generate the overall test score pattern.

Distinguishing the performance of Hispanic students by their generation helps shed light on the anomalous finding of steady gains relative to white and black peers. There are arguments for why Hispanics that have been in the U.S. for three or more generations might perform worse, the same, or better than white students. Most scholars believe that across generations there is regression towards the mean in most socioeconomic outcomes. Hispanics have been in the U.S. for fewer generations than whites on average, and it may be that third-generation and higher Hispanics today are the offspring of positively selected immigrant grandparents or great-grandparents. This type of selection would imply a moderately positive but constant Hispanic-white gap for later-generation Hispanics. On the other hand, Duncan and Trejo (2011) show that native-born Hispanics that self-identify as such tend to be negatively selected largely because more successful Hispanics are less likely to retain their Hispanic identity and report it in surveys. Although this kind of selection would give rise to negative gap, it would be no real economic cause for concern; a true measure of Hispanic ethnicity would show that these Hispanics track whites.⁴

There are also arguments for how rapidly and how completely immigrant children can assimilate. While we know that Hispanic students as a whole surpass the performance of similar white students in later grades, whether this stems from immigrant children surpassing white students in later grades has not been established. Although immigrant students surely come in at a disadvantage, this disadvantage might be small enough that rapid growth in their test scores could put them on par with whites in a few years. Any growth beyond this point

⁴To clarify, a true measure of Hispanic ethnicity would capture whether the student or any of his ancestors (parents, grandparents, etc.) was from Cuba, Mexico, Puerto Rico, South or Central America, or other Spanish culture.

must come from factors beyond cultural integration since full assimilation only implies that average native achievement and average immigrant achievement are equal. Alternatively, immigrant students could enter so far behind or have such a slow growth rate that I never observe complete convergence with whites during the ages in my sample. A slow convergence rate would be consistent with the slow rate of wage assimilation found in Borjas (1985); however, immigrant children likely catch up to natives faster than immigrant adults.

Given the overall upward trend in Hispanic test scores, I identify three of the more plausible scenarios and illustrate them in Figure 1. In Panel A, I show the case where Hispanic families across the board are better at improving their children's achievement scores, but immigrant Hispanics start at a lower level. The overall trendline represents an average of achievement levels but not of slopes. For the next two panels, I assume that Hispanics with native-born parents have flat trendlines relative to whites, as do black students. In Panel B, I depict the case where their achievement lies above whites due to positive selection of their immigrant grandparents and great-grandparents. To generate the overall trendline, recent immigrants must improve rapidly, but they would still score below whites in later grades. This scenario leaves open the possibility that the growth in Hispanic immigrant test scores tapers off as they approach whites. Finally, Panel C shows what happens when there is negative selection into ethnic identification for established Hispanic families. Their test scores then fall slightly below whites in each grade. To make up for the lower scores of later-generation immigrants still identifying as Hispanic, recent immigrants must assimilate quickly and go on to outscore whites by a wide margin. Then, the overall pattern would understate the relative performance of immigrant Hispanic youth in later grades as well as the speed of their convergence. In this final scenario, a factor besides assimilation must drive the growth in Hispanic test scores in these later grades.

2.2 Related empirical work

Despite the fact that Hispanics recently overtook blacks as the largest minority group, the Hispanic-white test score gap is almost always discussed as an afterthought to the black-white gap in papers on minority achievement. Outside of sample size concerns when Hispanics were a smaller part of the population, this feature of the literature on racial/ethnic inequality partly came about because of the difficulty in analyzing Hispanic-white differences due to heterogeneity among Hispanics. As they emerge as an important demographic group in their own right, researchers have begun to tackle the particular issues that Hispanics face.

Past studies usually estimate the raw Hispanic-white test score gap as smaller than the black-white gap but vary in how much smaller. For example, Phillips and Chin (2004) place it at 78% of the size of the black-white gap in math in 4th grade and 84% in reading using National Assessment of Educational Progress (NAEP) data. With a more recent wave of the same data set, Reardon and Galindo (2009) estimate the size of the Hispanic-white achievement gaps in reading and math to be about three quarters of a standard deviation in 4th and 8th grade. Unfortunately, the NAEP is more useful for studying trends across time than how Hispanic-white differences evolve as students progress through school since the data set contains a limited number of grades and samples repeated cross sections.

The Early Childhood Longitudinal Study-Kindergarten Class (ECLS-K) provides estimates on achievement gaps for some of the youngest ages. While this data is useful for analyzing some types of achievement inequality, language proficiency rules for administering assessments lower its value for studying high-immigrant populations.⁵ Nevertheless, Fryer and Levitt (2004, 2006) and Reardon and Galindo (2009) estimate Hispanic-white differences

⁵For one, students were not given the reading test if they were not proficient in oral English, meaning that there is severe sample selection into this assessment on the order of 29% for all Hispanics and 77% for first-generation Hispanics in the early waves of the survey. Also, students not proficient in oral English but proficient in oral Spanish were given the math assessment in Spanish. While these math scores do not suffer from the normal language bias, one cannot ascertain how improvement in English language skills contributes to any narrowing of the Hispanic-white math gap. Furthermore, some evidence indicates that a number of Hispanic students were misclassified as being proficient in oral English when parents later reported Spanish as the primary language spoken at home (Reardon and Galindo, 2009). Thus, one cannot be certain that math scores suffer from no language bias.

to bounce between -0.3 and -0.5 standard deviations in reading, with no clear trend. In math, they find that the raw gap in the fall of Kindergarten is around -0.8 standard deviations but shrinks to -0.5 standard deviations by the spring of 5th grade.

Crucial to the analysis of the evolution of the Hispanic-white gap for cohorts of students is whether the sample is restricted to continuous enrollees. Without this restriction, researchers find that the size of the gap is constant or even grows as recently arrived immigrants enter the sample. With it, they find progress. In addition, controlling for key background characteristics explains much of the Hispanic-white gap. Clotfelter, Ladd, and Vigdor (2009) condition on continuous enrollment in North Carolina public schools and find that between 3rd and 8th grade, Hispanic students close their reading and math test score gaps with similar whites in 5th grade and then finish 8th grade about 0.1 standard deviations ahead. Expanding on this paper, Clotfelter, Ladd, and Vigdor (2012) demonstrate that each group of late arrivers makes progress in the time it stays enrolled in the school system.

Past literature has also examined the impact of immigrant generation and time of arrival on a variety of educational outcomes, but there remains a lack of consensus on this topic as conclusions vary depending on the outcome studied and the racial/ethnic composition of the immigrant pool (see Kao and Tienda, 1995; Rong and Grant, 1992; Chiswick and DebBurman, 2004; Cortes, 2006). With the problematic ECLS-K data, Reardon and Galindo (2009) examine raw test score differences for Mexican students by immigrant generation. In math, they find the magnitude and pattern of raw gaps to be similar for first- and second-generation students. These students begin Kindergarten with math scores 1.1 standard deviations lower than whites and close the gap to around -0.8 standard deviations over the next two years, but it remains relatively constant through 5th grade. The third-generation Mexican students display a similar pattern, though they start school only 0.5 standard deviations below whites. The profile of reading scores resembles that of the math scores for the second and third generations; the authors do not report estimates for the first generation since so few were proficient enough in English to take the reading assessment.

The study most closely related to this one is a recent working paper by Özek and Figlio (2014). They conduct a similar analysis with the Florida analogue of the North Carolina data and come to some similar conclusions about how Hispanic students of various generations fare as they move through school. For the within-Hispanic comparisons, they find the same patterns as the ones presented here. They also find some convergence with whites for first- and second-generation Hispanics, but they do not show children of immigrants closing the achievement gap or surpassing whites. The discrepancy might arise because they do not control for parent's education while I do. This key control makes the white and Hispanic students under analysis more socioeconomically comparable. Another disadvantage of the Özek and Figlio (2014) study relative to this one is that the authors exclude all native-born Hispanic students born in a state besides Florida, or 23% of Hispanic students. In contrast, I use a finite mixture model to back out immigrant generation when it is unobserved and thus do not drop any observations. Still, the fact that data from two different states leads to some similar conclusions adds confidence to the generalizability of the findings presented in both papers.

This paper improves upon past work by breaking down trends in the Hispanic-white test score gap for a set of intact, complete cohorts. With this data set and econometric model, I establish which Hispanics are responsible for the trend reversal, the size of the test score gap for each generation, and how the gaps evolve as the students age. In the process, I determine how quickly test scores for immigrant Hispanics converge with whites of a similar socioeconomic status. By focusing on a single ethnic group, I acknowledge that the experience of immigrant Hispanics might be quite different from, for example, immigrant Asians.

3 Data

The North Carolina education data is a rich, longitudinal, administrative data set that links information on students, teachers, and public schools over time. The database is maintained by the North Carolina Education Research Data Center (NCERDC), which is housed at Duke University. Every year, the data center processes files received by the North Carolina Department of Public Instruction (NCDPI) and adds them to the existing data. Since the availability of some parts of the data varies over time, I restrict my analysis to certain years. I draw information on students from the End of Grade (EOG) files, which contain reading and math test scores for 3rd through 8th graders as well as student background characteristics such as race/ethnicity, parent's education, and free/reduced price lunch status. Encrypted identifiers follow students over time, even as they change schools, as long as they stay within the universe of North Carolina public schools.

My sample consists of two cohorts of students in 3rd grade in 2000 and 2001. For the 2000 cohort, normal grade progression means that these students would be in 4th grade in 2001, 5th grade in 2002, and so on, up to 8th grade in 2005 (and similarly for the 2001 cohort). Since late arrival into and early exit out of the sample are important variables in my analysis, I make some adjustments so that I do not confuse retention with entering or leaving the sample. Specifically, I follow students who were held back in 3rd grade or any subsequent grade, and I drop any students that enter a cohort in 4th grade or above due to grade repetition.⁶ This procedure ensures that my results are valid for the cohorts of students attending 3rd grade in 2000 and 2001 and that late arrival and early exit variables only capture students that enter and exit the universe of North Carolina public schools. Note

⁶With this procedure, I miss students that start attending North Carolina public schools at the same time that they repeat a grade. To check the extent to which this might be a problem, I also estimate my model restricting the sample to students who were born between October 1990 and September 1991 for the 2000 cohort and between October 1991 and September 1992 for the 2001 cohort. The cutoff for entering Kindergarten was October 16 for these cohorts, and the months in these ranges are also the most frequent months of birth, covering 77% of the sample. While the overall pattern of my results remains the same, I do find that all generations of Hispanic students perform worse relative to whites, as compared to the sample that does not drop students who were not the appropriate age for grade.

that for the rest of the paper “grade” means imputed grade, or the grade that a student would be in if he had been promoted each year. Finally, I normalize test scores within each cohort-grade pair.⁷ To maintain comparability across time given changes in the racial and ethnic composition of students, I use the mean and standard deviation of the white test score distribution in each cohort-grade pair for normalization.

The NCERDC has also matched students in the education files to birth records supplied by the North Carolina State Center for Health Statistics. These birth records provide even more background information on students, including the location of birth for each of the parents, but this information is not available for students born outside of North Carolina. Furthermore, the match rate of 3rd grade education data to birth data is 66% for all students but 19% for Hispanic students. I expect a lower match rate for Hispanics since first-generation immigrants do not have an in-state birth record by definition. Furthermore, some native-born Hispanics in North Carolina moved to the state between birth and the end of 3rd grade. If all Hispanic students born in North Carolina were matched, the education to birth data match rate would be 32%.⁸ I make use of the birth certificate information when possible and let my econometric model take care of the rest. When a student’s birth certificate indicates that at least one parent was born outside the country, I classify him as a second-generation immigrant.⁹ Students with both parents born in the U.S. are categorized as third-generation, but in reality, their roots could reach back even further.

For the group of Hispanic students without birth records, I rely on the Census and the American Community Survey (ACS) to fill in the distribution of immigrant generations for school-age Hispanic youth not born in North Carolina (Ruggles et al., 2010). To ensure the samples are comparable, I use school-age Hispanics living in North Carolina and attending public school from the 2000 Census and 2001-2006 waves of the ACS. I choose the 2000 and 2001 3rd grade cohorts so that the Census/ACS and NCERDC data line up in terms

⁷Test scores are reported on the same developmental scale regardless of grade, so I can include retained students in the normalization.

⁸Calculated from the Census/ACS. See Appendix A for the full details.

⁹This definition of second-generation is consistent with other social science research.

of available information. Earlier cohorts leave me without corresponding years in the Census/ACS. In later ones, the NCERDC does not provide free/reduced lunch status and parent education, the key variable to predict immigrant generation.

While the North Carolina education data is not nationally representative, there are several reasons that make the state interesting to use as a case study. First, North Carolina is in fact representative of the country in a few key ways. It features a diversified economy with a mixture of urban, suburban, and rural areas spread across the Appalachian Mountains, the Piedmont, and the Atlantic coast. Hispanics reside in all parts of the state, drawn to the demand for workers in such industries such as agriculture, construction, service, and retail (Gill, 2010). Second, while Hispanics make up a smaller proportion of North Carolina’s population than they do nationally (8.4% versus 16.3% as of 2010), the state has experienced astounding growth in the number of Hispanics residing there in the past two decades. Therefore, the state could serve as a model for what will happen in other parts of the nation.

4 Econometric model

If the immigrant generation for every Hispanic student was known, I could estimate generational test score gaps by ordinary least squares, taking the following equation as the model:

$$y_{ik} = g_i' \alpha_k + x_{ik}' \beta_k + \varepsilon_{ik} \tag{1}$$

where y_{ik} is the test score for student i in subject k , with $k = r$ for reading and $k = m$ for math; g_i is a vector of indicators for each Hispanic immigrant generation; and x_{ik} is a vector of other covariates, including indicators for the non-Hispanic racial groups. The parameter vector α_k then contains the Hispanic-white test score gap for each immigrant generation in subject k . However, I must take a different approach to uncover these values since since I do not observe generation for the majority of Hispanic students.

I estimate the effect of generational status on test scores by maximum likelihood with a

two-part likelihood function: The first component, which I use for observations where g_i is observed, comes from the standard likelihood for a linear model with a normally distributed error. The second component departs slightly from the standard likelihood to a finite mixture model since I must incorporate the fact that I do not observe immigrant generation for these students. In Appendix B, I show that estimates from my model are not substantially different from estimates obtained with ordinary least squares, as far as the overall Hispanic-white test score gap is concerned.

4.1 Likelihood function

In order to take full advantage of the information available, I jointly estimate achievement gaps for reading and math. This aspect of the setup is important for students with an unknown immigrant generation since it utilizes the correlation between the two scores; in other words, I use the fact that a high reading score may indicate that a student is of a higher immigrant generation in my estimation of the math coefficients, and vice versa. Specifically, I model the distribution of the shocks ε_{im} for math and ε_{ir} for reading as bivariate normal, where I parameterize the covariance matrix such that ρ is the correlation between the shocks and σ_k^2 is the variance of subject k 's shock.

$$\begin{bmatrix} \varepsilon_{ir} \\ \varepsilon_{im} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_r^2 & \rho\sigma_r\sigma_m \\ \rho\sigma_r\sigma_m & \sigma_m^2 \end{bmatrix} \right) \quad (2)$$

For a non-Hispanic student or a Hispanic student with a birth certificate and thus a known immigrant generation, the contribution to the likelihood function is

$$L_{1i} = \Phi(y_{ir}, y_{im} | g_i, x_{ir}, x_{im}, \theta) \quad (3)$$

where $\Phi(\cdot)$ is the bivariate normal probability density function, and θ contains all the pa-

rameters.¹⁰ For a Hispanic student with an unknown immigrant generation, the likelihood contribution is

$$L_{2i} = \sum_g \pi_{ig} \Phi(y_{ir}, y_{im} | g, x_{ir}, x_{im}, \theta) \quad (4)$$

where π_{ig} is the probability that student i belongs to generation g . In practice, this value is estimated in a first stage, as discussed below. Intuitively, this estimation method works because the objective function gives greater weight to observations that are likely to be of a particular immigrant generation when it determines the effect of being a member of that generation. The combined likelihood for all observations is

$$L = \prod_{i=1}^{n_1} L_{1i} \times \prod_{i=1}^{n_2} L_{2i} \quad (5)$$

Finally, taking logs and denoting the log-likelihood as ℓ , the function that I maximize to estimate the model is

$$\ell = \sum_{i=1}^{n_1} \ln \Phi(y_{ir}, y_{im} | g_i, x_{ir}, x_{im}, \theta) + \sum_{i=1}^{n_2} \ln \left[\sum_g \pi_{ig} \Phi(y_{ir}, y_{im} | g, x_{ir}, x_{im}, \theta) \right] \quad (6)$$

In practice, some math and reading scores are missing, and this issue is more prevalent for Hispanic students' reading scores since students with LEP (limited English proficient) status are exempt from the reading tests. (They are still required to take the math tests.) By incorporating information about whether or not a student took the other type of exam, I gain an even better hold on the immigrant generation to which a student belongs. Accordingly, I include a dummy in the vector x_{ik} for whether the student took the other type of exam. When a student is missing one exam but not the other, I replace the bivariate normal density

¹⁰That is,

$$\Phi(y_{ir}, y_{im} | g_i, x_{ir}, x_{im}, \theta) = \frac{1}{2\pi\sigma_r\sigma_m\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(y_{ir} - g'_i\alpha_r - x'_{ir}\beta_r)^2}{\sigma_r^2} + \frac{(y_{im} - g'_i\alpha_m - x'_{im}\beta_m)^2}{\sigma_m^2} - \frac{2\rho(y_{ir} - g'_i\alpha_r - x'_{ir}\beta_r)(y_{im} - g'_i\alpha_m - x'_{im}\beta_m)}{\sigma_r\sigma_m} \right] \right\}$$

with a univariate normal density for the test the student did take.

4.2 Estimation with school fixed effects

Due to computing and data constraints, I cannot estimate a model with school fixed effects using traditional methods (i.e., direct estimation of the school effects or de-meaning all variables using the within transformation). Therefore, I adopt the following iterative procedure.

1. Given initial guesses $(\hat{\alpha}_k, \hat{\beta}_k)$ for $k = \{r, m\}$, calculate the average residual for each school s in the set of schools:

$$\hat{\mu}_{sk} = \frac{1}{n_s} \sum_{i \in s} \hat{\varepsilon}_{ik} \quad (7)$$

$$\hat{\varepsilon}_{ik} = y_{ik} - g'_i \hat{\alpha}_k - x'_i \hat{\beta}_k \quad (8)$$

When g_i is not observed, I substitute the individual's immigrant generation probability distribution π_i . These average residuals by school represent the initial guesses for the school fixed effects.

2. Estimate the model with $(y_{ik} - \hat{\mu}_{sk})$ as the dependent variable to obtain new parameter guesses $(\hat{\alpha}_k, \hat{\beta}_k)$.
3. Repeat steps 1 and 2 until the estimated parameters converge.¹¹

4.3 First-stage estimation of probability weights

I estimate the weights π_{ig} using data from the Census and the ACS. Since both the education data and the Census/ACS contain parent education, I construct weights based on this characteristic. Specifically, I take the population of school-age Hispanic students born outside of but living North Carolina and attending public school. Then, I calculate the probability that a student is of each immigrant generation given parent's educational attainment in five

¹¹In practice, I stop iterating when the percent change in the log-likelihood is below a very small threshold.

categories.¹² Finally, I adjust the weights to correct for undermatching between the birth and education data. I explain this adjustment procedure and the reasoning behind it in more detail in Appendix A. While I prefer the weights based on parent’s education, unconditional weights yield the same qualitative results, which are also numerically very similar.¹³

5 Results

5.1 Descriptive statistics

Table 1 gives descriptive statistics for outcomes and covariates for the sample of students that are part of the cohorts in 3rd grade in 2000 and 2001. Pooling all grades together for each student gives over 1.2 million student-grade observations with an average of 4.7 grade observations per student, 3.7 for Hispanic students. In this sample, the average Hispanic-white reading gap is -0.74 standard deviations, and the average math gap is -0.61 standard deviations. The observation frequencies by grade in Table 2 justify normalization to the white mean and standard deviation for each cohort-grade combination—the number of white students in each grade is stable relative to the number of Hispanics. Due to this normalization, we would expect the year-to-year change in white test scores to be close to zero, though perhaps slightly greater than zero if there is positive selection into staying in the sample. For white math scores, this selection story appears to hold. However, the average Hispanic student gains about 0.06 standard deviations in each subject, conditional on having a score in the previous year. Given that there are fewer observations per student for Hispanics, the selection into having a previous year’s score is likely higher. Nevertheless, this growth gives us a sense of the gains Hispanic students make with continuous enrollment.

This sample is 59% white, 31% black, and 5% Hispanic, while the remainder are another

¹²For some years in my sample, I could also include location at the Super PUMA level in the calculation of the probability weights. This addition barely alters the results.

¹³Specifically, no generational coefficient moves by more than 0.06 standard deviations, and most are within 0.02 standard deviations in my adjusted gaps model.

(single) race or multiracial.¹⁴ Since Hispanic students move in and out of North Carolina public schools more often, the racial/ethnic composition looks slightly different without the grades pooled. This aspect of the data is also reflected in the proportion of student-grade observations that arrive late (defined as missing from the 3rd grade data) and exit early (missing from the 8th grade data). Most notably, 37% of Hispanic observations come from late arrivers, compared to 11% for whites, but the 13% probability that a Hispanic observation represents an early exiter is also significantly greater than the white probability. While white and Hispanic observations are equally likely to be missing a math score (0.1%), given that they have test score data, Hispanics are about 5 times as likely (0.4% versus 1.9%) not to have a reading score.

The two socioeconomic status variables starkly reveal the relative average disadvantage of Hispanic students. Though about 40% of both whites and Hispanics have high school graduate for parent's education, the distributions skew in different directions: Forty-seven percent of Hispanic student-grade observations have parent educational attainment of less than high school compared to 8% for whites. A white student is much more likely to have a parent who has completed some level of higher education. This disparity in socioeconomic status shows up again in the proportion of students who receive free or reduced price lunch, which is only 39% for white students but 90% for Hispanics.

In Table 3, I report summary statistics for Hispanic students in the NCERDC data based on what I know about the students' immigrant generation. I separate the students not matched to a birth certificate by whether they arrived in North Carolina public schools by 3rd grade. From the Census and ACS (see Table 5), we know that about half the Hispanic students not born in North Carolina are first-generation immigrants. Between the late and early arrivers to North Carolina, there is remarkably little difference in socioeconomic status as represented by parent's education and subsidized lunch status. However, the reading

¹⁴For the race/ethnicity variable, the respondent could only mark one of the following: American Indian, Asian, Black, Hispanic, Multiracial, White, or Other. I interpret black as non-Hispanic black and white as non-Hispanic white, though I cannot know definitively what the student, parent, or teacher thought when marking the child's race/ethnicity.

test scores of the late arrivers are half a standard deviation lower and the math scores 0.38 standard deviations lower than the early arrivers. Even the early arrivers score about half a standard deviation below whites in both subjects. Late arrivers gain 0.12 standard deviations in math and reading from the previous year, though the caveat of having a previous year's test score applies more strongly for these students. Still, this change is three times the average gain of early arrivers. Finally, the late arrivers are missing a reading score in 3.9% of cases, compared to 0.8% for early arrivers.

Turning to students with a known immigrant generation, the third-generation Hispanic students lag behind whites by about 0.4 standard deviations in reading and math. The second generation's respective gaps are -0.62 and -0.47 standard deviations. While the students I know to be third-generation have the same average change in test scores as whites, known second-generation Hispanic students improve 0.05 standard deviations relative to whites. Not only are test score levels different by immigrant generation and time of arrival, so are the slopes. I also observe differences in the socioeconomic status of these two groups, as far as I can tell from the birth record matches. These third-generation Hispanic students on average have parents with less education than white students, and these second-generation students in turn have parents with less education than the third generation. More parents of second-generation Hispanics did not finish high school than did. For free/reduced price lunch, all groups of Hispanics in Table 3 have high levels of enrollment in the program, but the known third generation enrolls 77% compared to over 90% for the known second-generation and unknown generation students.

The Census and ACS provide a more complete classification of the immigrant generation of school-age Hispanic youth living in North Carolina as well as more background characteristics. I present descriptive statistics from this data for the years corresponding to the education sample (2000-2006) in Table 4.¹⁵ Turning to average characteristics by immigrant

¹⁵I observe some differences between the reported parent's education distributions in the NCERDC data for Hispanic youth (compare the last column of Table 1 to the last column of Table 4). This disparity could arise from conditioning on observing a test score in the NCERDC sample, teacher reporting of parent's education, or the fact that in the Census/ACS sample, I only have parent's education for youth living with

generation, I generally find that a higher generational status means that a student comes from a more advantaged household. Average parent’s education, household income, and mother’s age at birth all suggest that third-generation Hispanic youth are positioned to do well in school. Exceptions are food stamp reciprocity and mother’s marital status, which appear to give first-generation children an edge. However, given household income, the low take-up rate of food stamps among families with first-generation Hispanic children likely reflects less familiarity with resources for low-income households.

Since first- and second-generation youth both have immigrant parents, one would expect their household characteristics to be similar but perhaps with higher socioeconomic status for households with parents that migrated before the birth of their child since they have resided in the U.S. for longer. Indeed, this pattern is close to what I find: Average household income is statistically no different, and the parent’s education distributions appear to be similar. Mother’s age at birth is understandably lower for first-generation children since some mothers migrate in the middle of their fertile years. Finally, over three quarters of first- and second-generation children come from Mexico with another 10% from Central America, whereas about half of third-generation children trace their origins to Mexico and a quarter to Puerto Rico.

5.2 Probability weights

Before estimating the test score model, I compute probability weights using the sample of school-age Hispanic youth attending public school and living in but not born in North Carolina. I then adjust the weights for undermatching using the procedure described in Appendix A. From Table 5, we see that there is substantial heterogeneity in generational status by parent’s education for these unmatched records. To determine how parent’s education helps predict immigrant generation, compare the probability of each parent education level to the overall probability in the last row of Table 5. For these students, higher educational

at least one parent.

attainment by a parent is generally associated with a higher generation.

5.3 Test score results

I lay out my test score results in a series of three sets of graphs, which correspond to different specifications of the test score model. Each graph plots the implied overall Hispanic-white test score gap by grade in addition to the gaps by immigrant generation. I present estimates for the implied gaps in Appendix B.

In Figure 2, I show the raw Hispanic-white test score gaps by grade for reading and math. The corresponding estimates and standard errors are in Tables 6 and 7. These unconditional differences reveal how far behind the average Hispanic student of a given immigrant generation is compared to the average white student. A few features of these graphs stand out: First, each generation comes closer to eliminating the Hispanic-white test score gap in reading and math, though the difference between generations varies by grade. Second, the first generation drives the downward trend in the overall Hispanic-white gap, so much so, in fact, that the overall trendline masks gains made by the second generation between 3rd and 8th grade. The third generation flatlines over this time period, showing no notable improvement or decline relative to whites. Clotfelter, Ladd, and Vigdor (2009) also find this modest decline in the overall Hispanic-white gap, but these graphs show more nuance.

Next, we turn to Figure 3, which plots raw gaps across grades for the population of students that enrolls continuously in school (estimates in Tables 8 and 9). To add these controls while still keeping the model tractable, I introduce a late arrival indicator for students without any 3rd grade test scores and an early exit indicator for students without any 8th grade test scores. Since I expect these effects to be different for first-generation Hispanics as late arrival more often than not means that the student just arrived in the U.S., I interact each of these indicators with first-generation Hispanic.¹⁶ Indeed, Figure 3 shows a reversal

¹⁶When I allow full interactions of late arrival and early exit with each race/ethnicity and Hispanic immi-

in the trendlines for the first generation, which also flips the downward trend in the implied overall gap. This finding mimics the reversal in the overall Hispanic-white test score gaps that Clotfelter, Ladd, and Vigdor (2009) find after restricting their sample to continuous enrollees.¹⁷ This paper shows that the immigrant students arriving in the schools drive the change, as opposed to all Hispanic students scoring lower after a move. In reading, first-generation Hispanics who arrive by 3rd grade and continuously enroll improve from -1.00 standard deviations to -0.54 standard deviations—an average increase of 0.08 standard deviations per year. Second-generation students gain a modest 0.06 standard deviations to finish 8th grade with a gap of -0.69 standard deviations. The slopes of the math trendlines are similar for these earlier generations, but the average sizes of the gaps are about 0.1 standard deviations smaller. In contrast, third-generation Hispanic students hover around 0.2 standard deviations below whites in reading and 0.3 standard deviations below in math. As a point of comparison, the raw black-white gap in both subjects is consistently around -0.8 standard deviations, almost always below the first generation’s gap with whites.

The coefficients on late arrival in Tables 8 and 9 show how this change takes place. We cannot place a purely causal interpretation on the late arrival and early exit coefficients; technically, these variables only represent whether a student was linked to an earlier or later test score record. Nevertheless, most of these students are moving in and out of the state, with some transfers in and out of private school or home school. From Table 4, we know that 72% of first-generation Hispanics that moved from out of state in the past year actually moved from another country. For reading, there is a relatively modest impact, usually around 0.1 standard deviations, for entering and leaving North Carolina public schools. However, the additional effect of late arrival for first-generation Hispanics is tremendous, ranging from -0.83 to -1.3 standard deviations. This group suffers no perceptible additional penalty for

grant generation, I do find that some of these interactions vary significantly by group. However, there is no clear pattern across grades outside of the main effects and first-generation interactions. Furthermore, only the interactions with first-generation Hispanic have a meaningful impact on the corresponding main effect (i.e., move the coefficient beyond the second decimal place). In other words, a complete set of interactions does not substantially change the second-generation and third-generation Hispanic coefficients.

¹⁷Comparable estimates for the years in my sample are in Appendix Table B.2.

exiting early. In math, all students experience some drop in test scores for arriving late and exiting early, and again this drop is larger for first-generation Hispanics. However, their additional penalty is smaller than in reading, ranging between -0.35 and -0.58 standard deviations. The more negative late arrival impacts in higher grades for the first generation drives the narrowing of their test score gap with whites.

These results give a more complete picture of why Clotfelter, Ladd, and Vigdor (2009) and others (Fryer and Levitt, 2004, 2006; Reardon and Galindo, 2009) find that a continuously followed sample of Hispanic students reduces the size of its test score gap with whites over time, while repeated cross sections, like in Phillips and Chin (2004), show a widening of the gap. Immigrant students that enter a longitudinal sample after the first wave, usually from out of the country, pull down the Hispanic average more each year. While others have speculated that this was the case, the evidence that I present here rules out the possibility that native-born Hispanics also make a significant contribution.

Last, I plot Hispanic-white test score gaps adjusted for parent's education in five categories, free/reduced lunch, gender, and school-by-year fixed effects as well as the late arrival and early exit indicators discussed above. We know from Table 1 that Hispanic and white students differ substantially in these measures of socioeconomic status. From Tables 3 and 4, we know that they also vary within the Hispanic population by immigrant generation. Finally, different generations of Hispanic students may attend different quality schools. Thus, Figure 4 (estimates in Tables 10 and 11) shows what differences remain between whites and Hispanic students of some immigrant generation after these additional controls are taken into account.

With this set of results, I find more support for the last scenario from Section 2.1, which I illustrate in Panel C of Figure 1. Third-generation Hispanics consistently perform about 0.1 standard deviations below whites in reading and math. If the Trejo critique holds, negative selection into Hispanic identification for later generations is responsible for this negative gap (Duncan and Trejo, 2011). The children of less successful Hispanics are on average more

likely to be classified as Hispanic at school, which leads to the estimation of slightly lower scores relative to whites for this group. For the children of immigrants, I find astounding progress in both subjects with continuous enrollment from 3rd to 8th grade. After a gap of -0.37 standard deviations in reading in 3rd grade, the scores of first-generation Hispanics are statistically no different than whites from 4th grade onward. In math, these students score 0.21 standard deviations below whites in 3rd grade, surpass them in 5th grade, and exit 8th grade 0.12 standard deviations ahead of them. For the average immigrant student that arrives by 3rd grade, assimilation in test scores is complete in just a few years. This pattern of growth not only applies to Hispanic students who are immigrants themselves but also to the native-born children of immigrants. Second-generation Hispanics start off in 3rd grade with scores on par with comparable whites in both subjects. In 8th grade, they outscore them by 0.11 standard deviations in reading and 0.24 standard deviations in math. Thus, these results show that the second generation more than assimilates; it outstrips whites as well as later-generation Hispanics.

Putting these three groups together, the trajectories of the implied overall gaps in math and reading appear very similar to those estimated in Clotfelter, Ladd, and Vigdor (2009). Between 3rd and 8th grade, the average Hispanic student that continuously enrolls reverses his position relative to the average white student from the same rough socioeconomic background, moving from 0.15 standard deviations below to 0.04 standard deviations above whites in reading, 0.11 below to 0.11 above whites in math. However, this overall trendline masks substantial heterogeneity in the growth and level of test scores by generational status.

The differences between the raw and adjusted test score plots come from adding controls for parent's education, free/reduced price lunch, gender, and school-by-year fixed effects. The magnitudes of the socioeconomic coefficients in Tables 10 and 11 reveal how important family background is to a child's performance in school. I estimate the effect of receiving a subsidized lunch as falling around -0.35 standard deviations, depending of the grade and subject. A student with a parent who graduated college scores almost a standard deviation

higher on exams than a student with a parent that dropped out of high school. Besides parental human capital, this measure of educational attainment may also capture different meanings of educational attainment in the U.S. versus abroad. For example, having a parent that did not finish high school carries a more negative association in the U.S. than in a country like Mexico, where a higher proportion of the population does not finish secondary school.¹⁸ The school-by-year fixed effects alone do not alter the estimated gaps by much, a finding consistent with other papers that use this data, including Clotfelter, Ladd, and Vigdor (2009).

With Florida data, Özek and Figlio (2014) also find growth in the test scores of the children of immigrants over the same range of grades, but their estimates do not show these students catching up with whites. There are a few differences between Özek and Figlio (2014) and this paper that could account for this discrepancy. First, Özek and Figlio (2014) exclude the 23% of Hispanics that were born inside the U.S. but outside of Florida. Households with the means to move between states most likely have a higher socioeconomic status, implying that the sample of native-born Hispanics in Özek and Figlio (2014) is negatively selected. This selection could explain why the second-generation Hispanic students in Özek and Figlio's paper only make slight progress relative to whites and never close the achievement gap. Second, Özek and Figlio (2014) do not control for parent's education. The large coefficients on the levels of parent's education in Tables 10 and 11, coupled with what we know about differences in the variable for white and Hispanic students from Table 1, show that its omission as a control is important. It means that the white and Hispanic students that Özek and Figlio (2014) compare in their regression-adjusted estimates are not as socioeconomically similar as they are in this paper. Relative to this paper, Özek and Figlio (2014) compare poorer Hispanic immigrants to whites, so it is unsurprising that they find a larger gap that

¹⁸To investigate this issue further, I estimate my model on the sample of students with a high school educated parent since this level was common among whites and Hispanics. The pattern of test score gaps for the first generation is similar, but I find that the trajectory of the second generation looks closer to that of the third. However, I may merely be picking up differences in time of arrival for the immigrant parents themselves. If an immigrant parent completed high school, it is more likely that the parent migrated as a child and thus was educated in the U.S.

does not close as students age. Last, there could be some underlying differences between the Hispanic and white populations in North Carolina and Florida that drive the partial disparity in results for the two papers. While I cannot rule out a role for this last factor, selection into the sample and the difference in controls seem likely culprits.

6 Conclusion

By decomposing the Hispanic-white test score gap by immigrant generation, this paper delves into why we see Hispanic test scores rise for intact cohorts and what it tells us about the academic progress of immigrant and native Hispanics. First, I show that the downward trend found with repeated cross sections becomes an upward trend after conditioning on whether first-generation students arrive late to the public school system. Previous work could not distinguish between this cause and all Hispanics suffering disproportionately after a move between states (Clotfelter, Ladd, and Vigdor, 2009, 2012).

Second, I establish where the performance of later-generation Hispanics lies relative to whites. Just as is true for black students, Hispanic students with U.S.-born parents score below whites in early grades, and there is no evidence of improvement or deterioration in their relative position as they age through school. Furthermore, the small negative gap between later-generation Hispanics and whites is consistent with Duncan and Trejo's (2011) evidence that the later-generation Hispanics who identify as such are negatively selected. If this selection story is true, their test score gap is no real economic cause for concern; this difference would not exist with a true measure of ethnicity.

Lastly and most importantly, I show that the overall gains in Hispanic test scores are driven entirely by students with immigrant parents and that convergence in test scores is quite rapid. These students begin school with scores below their white peers but by 8th grade often perform at substantially higher levels than them in both reading and math. The speed of assimilation in student test scores stands in contrast to the slower speed of wage

assimilation for adults and is consistent with greater adaptability among younger immigrants (Borjas, 1985). The findings of Clotfelter, Ladd, and Vigdor (2009) therefore understate the impressive progress of children with immigrant parents with years in school. That immigrant students eventually outscore whites calls the importance of assimilation into question. Although factors like language acquisition surely contribute to test score growth for immigrants at early ages, assimilation does not explain gains after parity with native whites is reached. Moreover, native-born children with immigrant parents have contact with U.S. institutions from birth, which also goes against the notion that U.S. experience alone accounts for immigrant children's progress. One explanation that is consistent with all the evidence is that children with immigrant parents have home environments that foster achievement, and this upbringing allows them to succeed over time. In any case, the progress made by Hispanic children of immigrants must come from a combination of noncognitive ability and home environment since I control for school environment and Cunha et al. (2006) rule out changes in cognitive ability in this age range.

The main takeaway is that Hispanic students from an immigrant background fare quite well under the current system, even better than suggested by Clotfelter, Ladd, and Vigdor (2009). Their main disadvantage is socioeconomic. While this hardship should not be downplayed—it still leads to the large raw achievement gaps between whites and immigrant Hispanics—the good news is that effective interventions targeted toward poor students should lift up Hispanic achievement as well.

References

- Borjas, George J. 1985. "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* 3 (4):463–489.
- Chiswick, Barry R. and Noyna DebBurman. 2004. "Educational attainment: analysis by immigrant generation." *Economics of Education Review* 23:361–79.
- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2009. "The Academic Achievement Gap in Grades 3 to 8." *Review of Economics and Statistics* 91 (2):398–419.
- . 2012. "New Destinations, New Trajectories? The Educational Progress of Hispanic Youth in North Carolina." *Child Development* 83 (5):1608–22.
- Cortes, Kalena. 2006. "The effects of age at arrival and enclave schools on the academic performance of immigrant children." *Economics of Education Review* 25:121–35.
- Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, vol. 1, edited by Eric Hanushek and Finis Welch. Amsterdam: Elsevier, 697–812.
- Duncan, Brian and Stephen J. Trejo. 2011. "Intermarriage and the Intergenerational Transmission of Ethnic Identity and Human Capital for Mexican Americans." *Journal of Labor Economics* 29 (2):195–227.
- Fryer, Roland G., Jr. and Steven D. Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School." *Review of Economics and Statistics* 86 (2):447–464.
- . 2006. "The Black-White Test Score Gap Through Third Grade." *American Law and Economics Review* 8 (2):249–281.
- Gill, Hannah. 2010. *The Latino Migration Experience in North Carolina: New Roots in the Old North State*. Chapel Hill, NC: The University of North Carolina State Press.

- Kao, Grace and Marta Tienda. 1995. "Optimism and Achievement: The Educational Performance of Immigrant Youth." *Social Science Quarterly* 76 (1):1–19.
- Özek, Umut and David N. Figlio. 2014. "Immigrants in Public Schools: A Closer Look at Cross-Generational Differences in Student Outcomes." Working Paper.
- Phillips, Meredith and Tiffani Chin. 2004. "School Inequality: What Do We Know?" In *Social Inequality*, edited by Kathryn Neckerman. New York: Russel Sage Foundation.
- Reardon, Sean F. and Claudia Galindo. 2009. "The Hispanic-White Achievement Gap in Math and Reading in the Elementary Grades." *American Educational Research Journal* 46 (3):853–891.
- Rong, Xue Lan and Linda Grant. 1992. "Ethnicity, Generation, and School Attainment of Asians, Hispanics, and Non-Hispanic Whites." *Sociological Quarterly* 33 (4):625–36.
- Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]*. Minneapolis: University of Minnesota.

Table 1: Descriptive statistics, NCERDC data

	All students	White	Hispanic
Reading score	-0.307 (0.001)	0.000 (0.001)	-0.743 (0.004)
Change in reading score	0.001 (0.001)	0.000 (0.001)	0.062 (0.003)
Math score	-0.301 (0.001)	0.000 (0.001)	-0.613 (0.004)
Change in math score	0.011 (0.001)	0.004 (0.001)	0.063 (0.003)
<i>Race/ethnicity</i>			
White	0.593		
Black	0.308		
Hispanic	0.050		
Other race	0.032		
Multiracial	0.018		
<i>Parent's education</i>			
Less than high school	0.124	0.081	0.467
High school graduate	0.462	0.423	0.412
Junior college or trade school	0.166	0.184	0.056
Four-year college	0.203	0.272	0.054
Graduate school	0.041	0.057	0.011
Free/reduced lunch	0.573	0.386	0.898
Female	0.492	0.490	0.495
Missing reading score	0.006	0.004	0.019
Missing math score	0.001	0.001	0.001
Late arrival	0.132	0.113	0.374
Early exit	0.092	0.091	0.127
Student-grade observations	1,257,894	745,796	63,365

Standard errors in parentheses. Sample: pooled grades 3-8 of the cohorts in 3rd grade in 2000 and 2001.

Table 2: Sample sizes by cohort and grade, NCERDC data

	All students		White		Hispanic	
	2000	2001	2000	2001	2000	2001
3rd	100,852	101,963	61,982	60,934	3,212	4,542
4th	101,303	102,481	61,821	60,954	3,668	4,964
5th	102,243	105,344	62,005	61,755	4,179	5,766
6th	105,046	106,953	62,628	62,111	4,932	6,330
7th	107,075	108,199	63,235	62,386	5,598	6,848
8th	107,897	108,538	63,540	62,445	6,072	7,245
Total	1,257,894		745,796		63,356	

Sample: cohorts in 3rd grade in 2000 and 2001.

Table 3: Descriptive statistics, Hispanics in the NCERDC data

	Unknown generation		Known generation	
	Late arriver	Early arriver	2nd gen	3rd gen
Reading score	-1.012 (0.008)	-0.507 (0.006)	-0.621 (0.013)	-0.413 (0.016)
Change in reading score	0.119 (0.006)	0.043 (0.004)	0.044 (0.009)	0.003 (0.011)
Math score	-0.862 (0.006)	-0.482 (0.005)	-0.468 (0.012)	-0.400 (0.015)
Change in math score	0.122 (0.005)	0.040 (0.003)	0.051 (0.008)	0.004 (0.010)
<i>Parent's education</i>				
Less than high school	0.465	0.475	0.584	0.249
High school graduate	0.417	0.406	0.350	0.518
Junior college or trade school	0.050	0.056	0.038	0.117
Four-year college	0.057	0.052	0.024	0.099
Graduate school	0.011	0.012	0.005	0.018
Free/reduced lunch	0.899	0.908	0.926	0.769
Female	0.479	0.500	0.519	0.516
Missing reading score	0.039	0.008	0.007	0.005
Missing math score	0.001	0.001	0.000	0.001
Late arrival	1.000	0.000	0.082	0.097
Early exit	0.145	0.133	0.048	0.079
Student-grade observations	22,881	31,147	5,490	3,838

Standard errors in parentheses. Sample: pooled grades 3-8 of the cohorts in 3rd grade in 2000 and 2001.

Table 4: Descriptive statistics, Hispanics in the Census and ACS data

	1st gen	2nd gen	3+ gen	All Hispanics
<i>Parent's education</i>				
Less than high school	0.497	0.578	0.152	0.432
High school graduate	0.306	0.327	0.494	0.367
Junior college or trade school	0.054	0.038	0.148	0.074
Four-year college	0.079	0.041	0.143	0.082
Graduate school	0.065	0.017	0.063	0.045
Household income (\$1000s)	40.9 (1.8)	37.6 (1.6)	53.9 (2.5)	42.9 (1.1)
Food stamp reciprocity	0.109	0.191	0.186	0.160
Mother's age at birth	24.2 (0.4)	25.3 (0.2)	26.1 (0.3)	25.2 (0.2)
Mother married	0.810	0.716	0.718	0.748
<i>Mother's employment status</i>				
Employed	0.509	0.505	0.607	0.536
Unemployed	0.083	0.083	0.099	0.088
Not in labor force	0.408	0.412	0.294	0.377
Missing mother's characteristics	0.213	0.119	0.054	0.136
<i>Migration status, one year</i>				
Same house	0.701	0.759	0.776	0.742
Moved within state	0.179	0.164	0.156	0.167
Moved between states	0.034	0.063	0.060	0.051
Abroad one year ago	0.087	0.014	0.008	0.039
<i>Place of origin</i>				
Mexico	0.760	0.753	0.479	0.683
Central America	0.096	0.100	0.050	0.085
South America	0.065	0.022	0.049	0.045
Puerto Rico	0.036	0.040	0.285	0.104
Cuba	0.015	0.008	0.019	0.013
Other/not specified ^a	0.027	0.078	0.118	0.070
Maximum observations	1175	1225	1024	3424

Standard errors in parentheses. Sample: school-age Hispanic youth living in North Carolina and attending public school, 2000-2006. Used person-level weights provided by IPUMS.

^a Other countries represented are Spain and the Dominican Republic, but most of these respondents did not specify a particular country or region.

Table 5: Probability weights by parent's education

	1st gen	2nd gen	3+ gen
Less than high school	0.461	0.450	0.089
High school graduate	0.324	0.329	0.347
Junior college or trade school	0.269	0.285	0.446
Four-year college	0.367	0.175	0.458
Graduate school	0.513	0.146	0.341
All	0.452	0.332	0.216

Sample: school-age Hispanic youth attending public school and living in but not born in North Carolina, 2000-2006. The person-level weights provided by IPUMS are adjusted to account for unmatched education records (see text for details).

Table 6: Raw reading gaps by grade

Grade	3rd	4th	5th	6th	7th	8th
1st generation Hispanic	-1.017 (0.050)	-1.016 (0.043)	-1.235 (0.033)	-1.266 (0.029)	-1.429 (0.025)	-1.563 (0.023)
2nd generation Hispanic	-0.742 (0.035)	-0.732 (0.035)	-0.666 (0.025)	-0.689 (0.027)	-0.557 (0.022)	-0.597 (0.022)
3+ generation Hispanic	-0.213 (0.027)	-0.216 (0.028)	-0.227 (0.027)	-0.238 (0.023)	-0.194 (0.023)	-0.220 (0.023)
Black	-0.780 (0.005)	-0.822 (0.005)	-0.830 (0.005)	-0.840 (0.005)	-0.837 (0.005)	-0.848 (0.005)
Other race	-0.368 (0.013)	-0.369 (0.013)	-0.427 (0.013)	-0.378 (0.012)	-0.340 (0.013)	-0.332 (0.012)
Multiracial	-0.240 (0.018)	-0.230 (0.017)	-0.230 (0.019)	-0.233 (0.019)	-0.241 (0.017)	-0.225 (0.016)
Constant	0.000 (0.003)	0.000 (0.003)	0.000 (0.003)	0.001 (0.003)	0.000 (0.003)	0.001 (0.003)
Missing math score	-1.186 (0.100)	-1.037 (0.082)	-1.207 (0.069)	-1.463 (0.084)	-1.323 (0.060)	-1.388 (0.056)

Standard errors in parentheses. Sample: cohorts in 3rd grade in 2000 and 2001.

Table 7: Raw math gaps by grade

Grade	3rd	4th	5th	6th	7th	8th
1st generation Hispanic	-0.765 (0.035)	-0.751 (0.030)	-0.819 (0.028)	-0.935 (0.028)	-0.942 (0.021)	-0.930 (0.020)
2nd generation Hispanic	-0.662 (0.027)	-0.613 (0.024)	-0.587 (0.023)	-0.572 (0.027)	-0.583 (0.020)	-0.563 (0.020)
3+ generation Hispanic	-0.286 (0.028)	-0.256 (0.026)	-0.270 (0.028)	-0.259 (0.024)	-0.291 (0.023)	-0.273 (0.024)
Black	-0.876 (0.005)	-0.831 (0.005)	-0.841 (0.005)	-0.849 (0.005)	-0.816 (0.005)	-0.832 (0.005)
Other race	-0.280 (0.013)	-0.249 (0.013)	-0.234 (0.012)	-0.175 (0.011)	-0.129 (0.012)	-0.094 (0.012)
Multiracial	-0.293 (0.018)	-0.276 (0.016)	-0.279 (0.017)	-0.275 (0.018)	-0.313 (0.016)	-0.298 (0.016)
Constant	0.005 (0.003)	0.006 (0.003)	0.005 (0.003)	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)
Missing reading score	-0.851 (0.026)	-0.861 (0.022)	-1.006 (0.027)	-1.242 (0.035)	-1.072 (0.031)	-1.105 (0.034)

Standard errors in parentheses. Sample: cohorts in 3rd grade in 2000 and 2001.

Table 8: Raw reading gaps by grade with continuous enrollment

Grade	3rd	4th	5th	6th	7th	8th
1st generation Hispanic	-0.998 (0.033)	-0.701 (0.043)	-0.662 (0.120)	-0.591 (0.045)	-0.527 (0.112)	-0.544 (0.038)
2nd generation Hispanic	-0.747 (0.029)	-0.776 (0.029)	-0.753 (0.057)	-0.738 (0.027)	-0.676 (0.028)	-0.686 (0.022)
3+ generation Hispanic	-0.206 (0.026)	-0.221 (0.027)	-0.246 (0.043)	-0.231 (0.024)	-0.192 (0.023)	-0.239 (0.026)
Late arrival		-0.094 (0.009)	-0.069 (0.009)	-0.111 (0.006)	-0.077 (0.007)	-0.081 (0.006)
Early exit	-0.111 (0.007)	-0.103 (0.007)	-0.150 (0.008)	-0.286 (0.008)	-0.416 (0.009)	
Late arrival X 1st gen Hispanic		-0.849 (0.064)	-1.051 (0.144)	-0.994 (0.057)	-1.140 (0.087)	-1.312 (0.048)
Early exit X 1st gen Hispanic	0.014 (0.059)	-0.015 (0.079)	0.210 (0.084)	0.103 (0.062)	0.095 (0.152)	
Black	-0.782 (0.005)	-0.823 (0.005)	-0.831 (0.005)	-0.837 (0.005)	-0.832 (0.005)	-0.847 (0.005)
Other race	-0.366 (0.014)	-0.364 (0.013)	-0.419 (0.012)	-0.366 (0.013)	-0.329 (0.012)	-0.325 (0.013)
Multiracial	-0.233 (0.018)	-0.219 (0.018)	-0.218 (0.017)	-0.213 (0.017)	-0.219 (0.017)	-0.215 (0.017)
Constant	0.019 (0.003)	0.019 (0.003)	0.023 (0.003)	0.039 (0.003)	0.034 (0.003)	0.016 (0.003)
Missing math score	-1.168 (0.147)	-0.987 (0.062)	-1.171 (0.065)	-1.351 (0.074)	-1.170 (0.069)	-1.370 (0.051)

Standard errors in parentheses. Sample: cohorts in 3rd grade in 2000 and 2001.

Table 9: Raw math gaps by grade with continuous enrollment

Grade	3rd	4th	5th	6th	7th	8th
1st generation Hispanic	-0.753 (0.032)	-0.613 (0.034)	-0.504 (0.075)	-0.511 (0.044)	-0.509 (0.096)	-0.437 (0.037)
2nd generation Hispanic	-0.666 (0.031)	-0.605 (0.022)	-0.587 (0.036)	-0.560 (0.027)	-0.587 (0.021)	-0.564 (0.020)
3+ generation Hispanic	-0.278 (0.025)	-0.238 (0.027)	-0.259 (0.044)	-0.225 (0.024)	-0.263 (0.027)	-0.264 (0.026)
Late arrival		-0.164 (0.009)	-0.150 (0.008)	-0.167 (0.006)	-0.130 (0.007)	-0.127 (0.005)
Early exit	-0.128 (0.006)	-0.122 (0.006)	-0.163 (0.007)	-0.323 (0.008)	-0.400 (0.009)	
Late arrival X 1st gen Hispanic		-0.346 (0.059)	-0.533 (0.088)	-0.571 (0.051)	-0.514 (0.067)	-0.583 (0.044)
Early exit X 1st gen Hispanic	0.057 (0.062)	0.030 (0.066)	0.118 (0.066)	0.125 (0.068)	0.187 (0.105)	
Black	-0.876 (0.005)	-0.837 (0.005)	-0.819 (0.005)	-0.846 (0.005)	-0.807 (0.005)	-0.838 (0.005)
Other race	-0.277 (0.014)	-0.242 (0.013)	-0.221 (0.012)	-0.159 (0.012)	-0.114 (0.012)	-0.083 (0.013)
Multiracial	-0.286 (0.018)	-0.0260 (0.017)	-0.261 (0.017)	-0.249 (0.016)	-0.286 (0.015)	-0.282 (0.017)
Constant	0.027 (0.003)	0.032 (0.003)	0.039 (0.003)	0.052 (0.003)	0.043 (0.003)	0.025 (0.003)
Missing reading score	-0.836 (0.025)	-0.782 (0.022)	-0.973 (0.026)	-1.123 (0.037)	-0.905 (0.033)	-1.043 (0.039)

Standard errors in parentheses. Sample: cohorts in 3rd grade in 2000 and 2001.

Table 10: Adjusted reading gaps by grade with continuous enrollment

Grade	3rd	4th	5th	6th	7th	8th
1st generation Hispanic	-0.366 (0.048)	-0.066 (0.046)	-0.009 (0.046)	-0.011 (0.033)	0.064 (0.044)	0.047 (0.038)
2nd generation Hispanic	-0.028 (0.030)	-0.051 (0.030)	0.002 (0.027)	0.066 (0.021)	0.126 (0.023)	0.108 (0.020)
3+ generation Hispanic	-0.078 (0.032)	-0.089 (0.030)	-0.119 (0.029)	-0.095 (0.026)	-0.060 (0.026)	-0.088 (0.024)
Late arrival		-0.130 (0.008)	-0.118 (0.006)	-0.188 (0.005)	-0.149 (0.005)	-0.155 (0.005)
Early exit	-0.118 (0.005)	-0.102 (0.006)	-0.131 (0.007)	-0.204 (0.007)	-0.287 (0.009)	
Late arrival X 1st gen Hispanic		-0.809 (0.069)	-1.043 (0.057)	-0.974 (0.038)	-1.067 (0.051)	-1.271 (0.045)
Early exit X 1st gen Hispanic	-0.106 (0.068)	-0.077 (0.074)	0.188 (0.068)	-0.029 (0.062)	-0.094 (0.072)	
Black	-0.432 (0.005)	-0.456 (0.005)	-0.454 (0.005)	-0.444 (0.005)	-0.449 (0.005)	-0.458 (0.005)
Other race	-0.137 (0.012)	-0.089 (0.011)	-0.113 (0.011)	-0.093 (0.011)	-0.056 (0.011)	-0.056 (0.011)
Multiracial	-0.074 (0.015)	-0.060 (0.015)	-0.052 (0.015)	-0.044 (0.014)	-0.052 (0.015)	-0.042 (0.015)
Constant	-0.568 (0.007)	-0.519 (0.007)	-0.536 (0.008)	-0.500 (0.007)	-0.523 (0.007)	-0.525 (0.007)
Missing math score	-0.952 (0.108)	-0.853 (0.073)	-0.908 (0.058)	-1.073 (0.063)	-0.907 (0.059)	-1.023 (0.050)
<i>Parent's education</i>						
High school graduate	0.414 (0.006)	0.371 (0.006)	0.390 (0.007)	0.347 (0.006)	0.373 (0.006)	0.372 (0.006)
Jr. college/trade school	0.656 (0.008)	0.603 (0.008)	0.635 (0.008)	0.575 (0.008)	0.603 (0.008)	0.599 (0.008)
Four-year college	0.956 (0.008)	0.915 (0.008)	0.920 (0.008)	0.868 (0.008)	0.896 (0.008)	0.886 (0.008)
Graduate school	1.206 (0.012)	1.173 (0.012)	1.171 (0.012)	1.122 (0.011)	1.145 (0.012)	1.142 (0.011)
Free/reduced lunch	-0.326 (0.005)	-0.340 (0.005)	-0.345 (0.005)	-0.335 (0.005)	-0.351 (0.005)	-0.356 (0.005)
Female	0.188 (0.004)	0.167 (0.004)	0.174 (0.004)	0.205 (0.004)	0.211 (0.004)	0.190 (0.004)
School X year FEs	yes	yes	yes	yes	yes	yes

Standard errors in parentheses. Sample: cohorts in 3rd grade in 2000 and 2001.

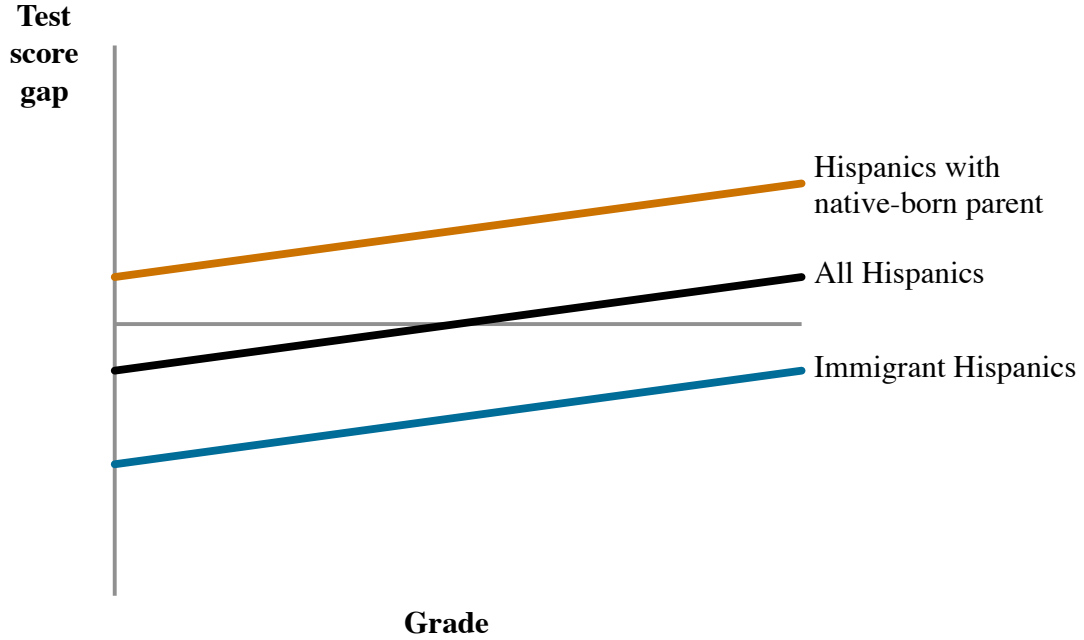
Table 11: Adjusted math gaps by grade with continuous enrollment

Grade	3rd	4th	5th	6th	7th	8th
1st generation Hispanic	-0.205 (0.039)	-0.053 (0.036)	0.043 (0.035)	0.061 (0.033)	0.070 (0.034)	0.120 (0.035)
2nd generation Hispanic	-0.006 (0.026)	0.046 (0.024)	0.111 (0.022)	0.241 (0.021)	0.165 (0.019)	0.235 (0.019)
3+ generation Hispanic	-0.152 (0.030)	-0.102 (0.027)	-0.126 (0.026)	-0.100 (0.025)	-0.100 (0.023)	-0.113 (0.023)
Late arrival		-0.199 (0.008)	-0.204 (0.006)	-0.247 (0.005)	-0.204 (0.005)	-0.205 (0.005)
Early exit	-0.141 (0.005)	-0.122 (0.006)	-0.152 (0.006)	-0.239 (0.007)	-0.278 (0.008)	
Late arrival X 1st gen Hispanic		-0.339 (0.058)	-0.533 (0.047)	-0.591 (0.036)	-0.486 (0.040)	-0.544 (0.039)
Early exit X 1st gen Hispanic	-0.039 (0.065)	-0.029 (0.060)	0.113 (0.058)	0.036 (0.053)	0.020 (0.056)	
Black	-0.543 (0.005)	-0.500 (0.005)	0.476 (0.005)	-0.439 (0.004)	-0.419 (0.004)	-0.424 (0.004)
Other race	-0.065 (0.011)	0.029 (0.011)	0.093 (0.011)	0.151 (0.011)	0.164 (0.010)	0.201 (0.010)
Multiracial	-0.135 (0.015)	-0.107 (0.014)	-0.099 (0.014)	-0.065 (0.014)	-0.099 (0.013)	-0.088 (0.014)
Constant	-0.417 (0.007)	-0.374 (0.007)	-0.362 (0.007)	-0.387 (0.007)	-0.381 (0.007)	-0.412 (0.007)
Missing reading score	-0.627 (0.022)	-0.599 (0.019)	-0.738 (0.022)	-0.822 (0.030)	-0.696 (0.029)	-0.843 (0.030)
<i>Parent's education</i>						
High school graduate	0.358 (0.006)	0.309 (0.006)	0.296 (0.006)	0.302 (0.006)	0.294 (0.006)	0.301 (0.006)
Jr. college/trade school	0.579 (0.008)	0.518 (0.007)	0.516 (0.007)	0.523 (0.007)	0.501 (0.007)	0.517 (0.007)
Four-year college	0.894 (0.008)	0.852 (0.007)	0.845 (0.008)	0.845 (0.007)	0.847 (0.007)	0.847 (0.007)
Graduate school	1.163 (0.012)	1.135 (0.011)	1.134 (0.011)	1.117 (0.011)	1.146 (0.011)	1.118 (0.011)
Free/reduced lunch	-0.307 (0.005)	-0.316 (0.005)	-0.323 (0.005)	-0.336 (0.005)	-0.346 (0.005)	-0.346 (0.005)
Female	0.003 (0.004)	0.013 (0.004)	0.019 (0.004)	0.063 (0.004)	0.066 (0.004)	0.070 (0.004)
School X year FEs	yes	yes	yes	yes	yes	yes

Standard errors in parentheses. Sample: cohorts in 3rd grade in 2000 and 2001.

Figure 1: Three scenarios generating the overall Hispanic trend

(A) Same slope, different intercepts



(B) Positively selected immigrant ancestors

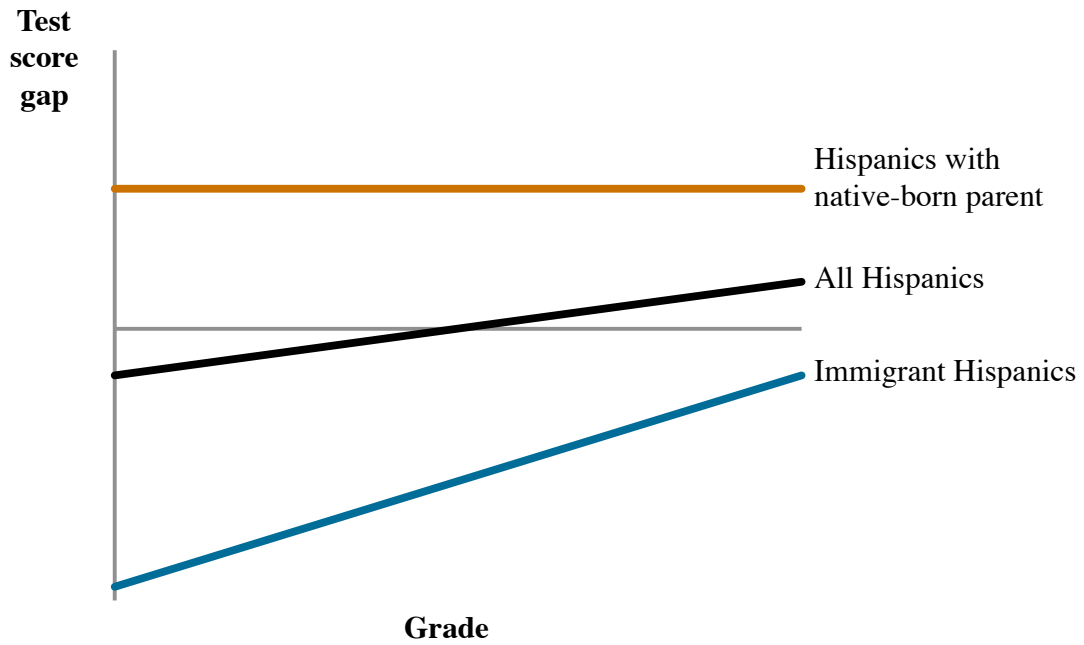


Figure 1: Three scenarios generating the overall Hispanic trend

(C) Negative selection into ethnic identification

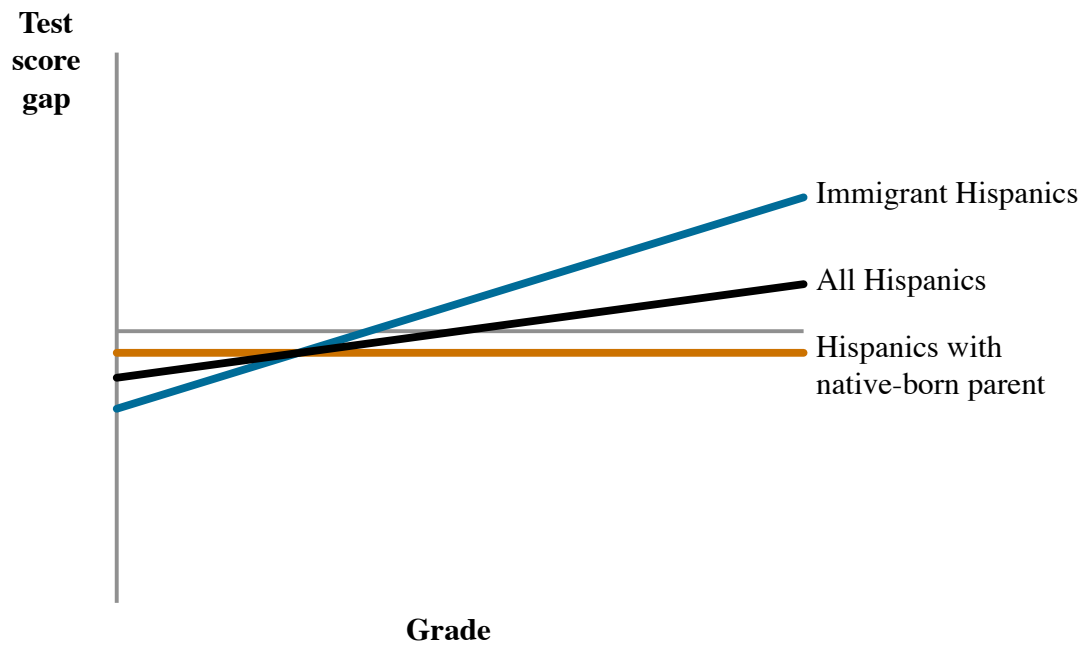
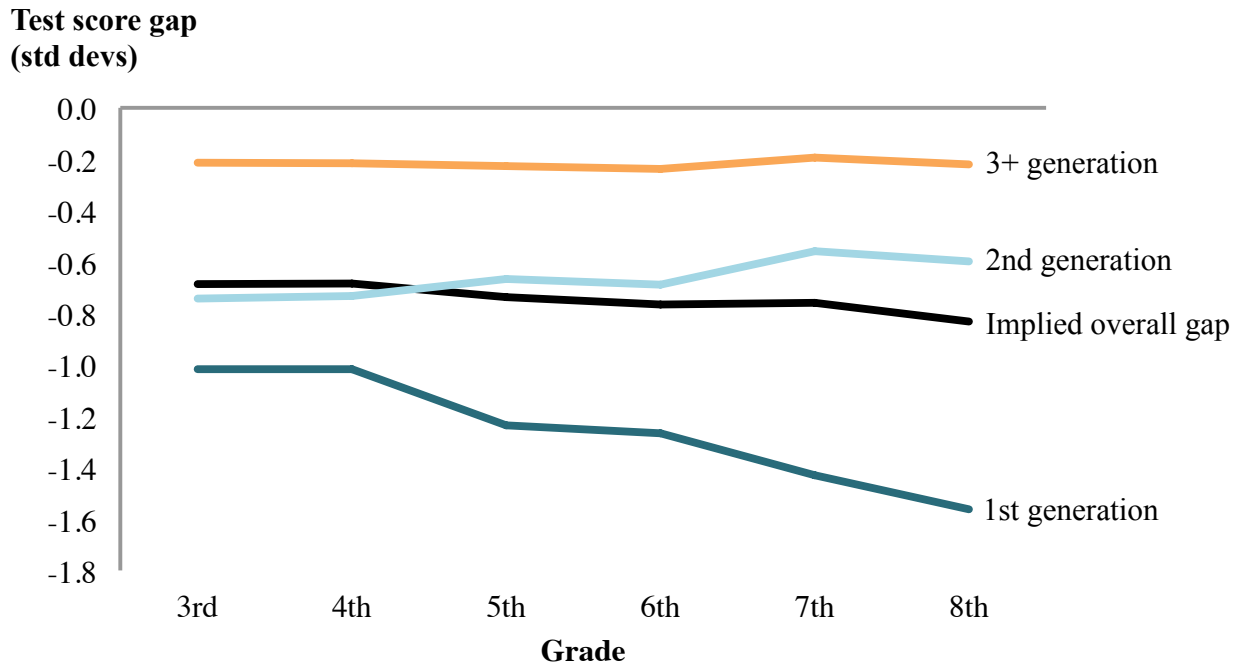


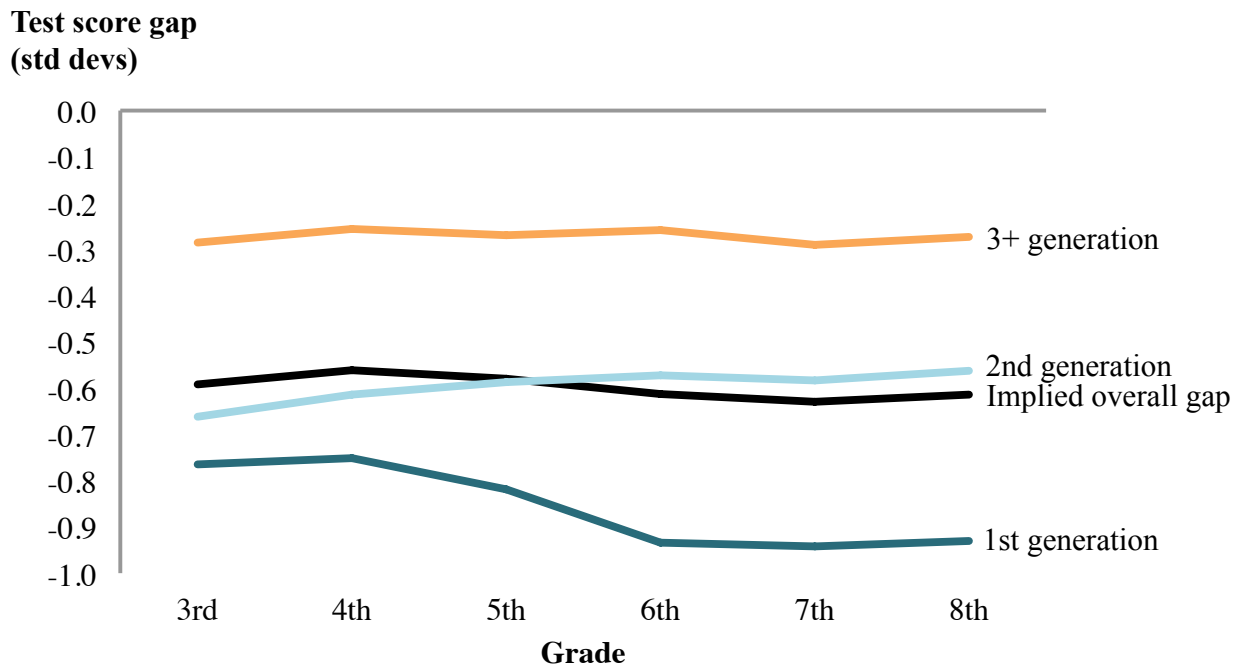
Figure provided for illustrative purposes only.

Figure 2: Raw Hispanic-white test score gaps by grade

(A) Reading



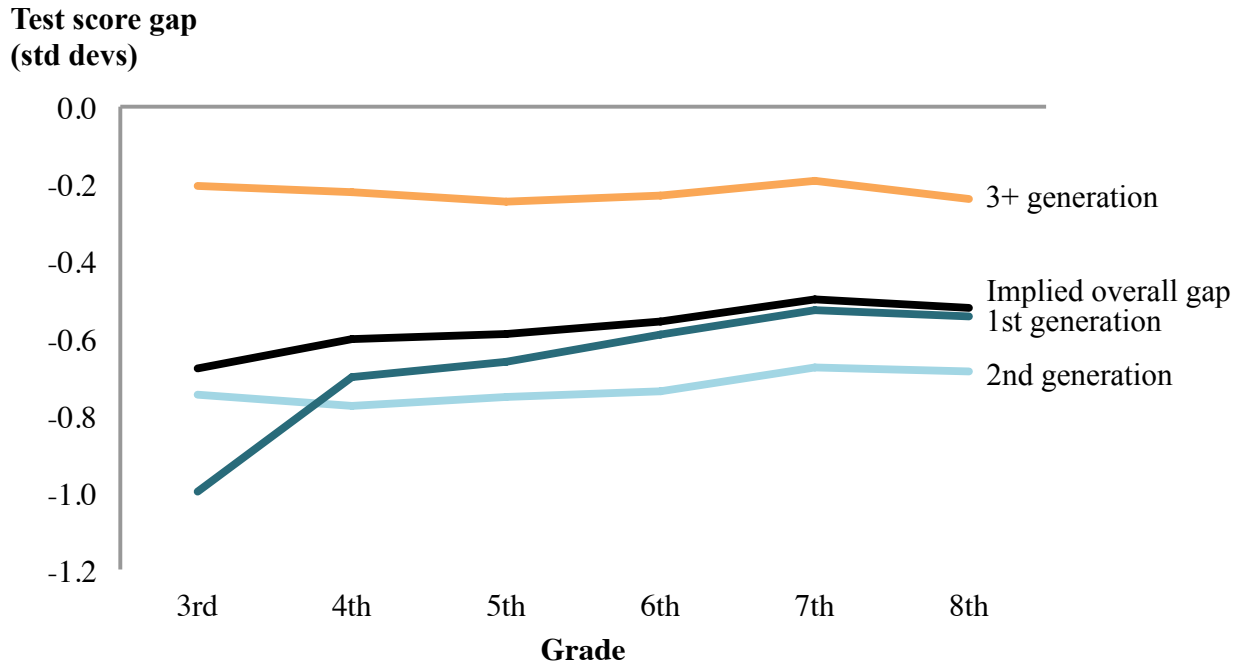
(B) Math



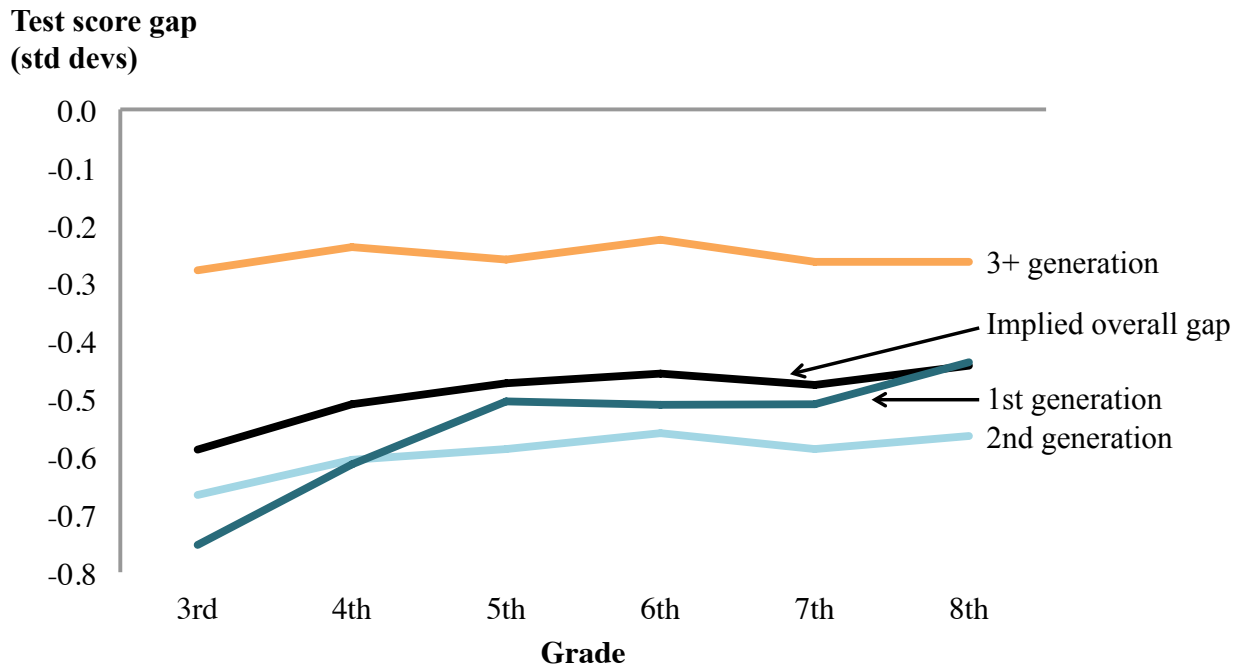
Sample: cohorts in 3rd grade in 2000 and 2001. Regressions also include indicators for black, other race, multiracial, and missing other test score. See Tables 6 and 7 for point estimates and standard errors.

Figure 3: Raw Hispanic-white test score gaps by grade for continuous enrollees

(A) Reading



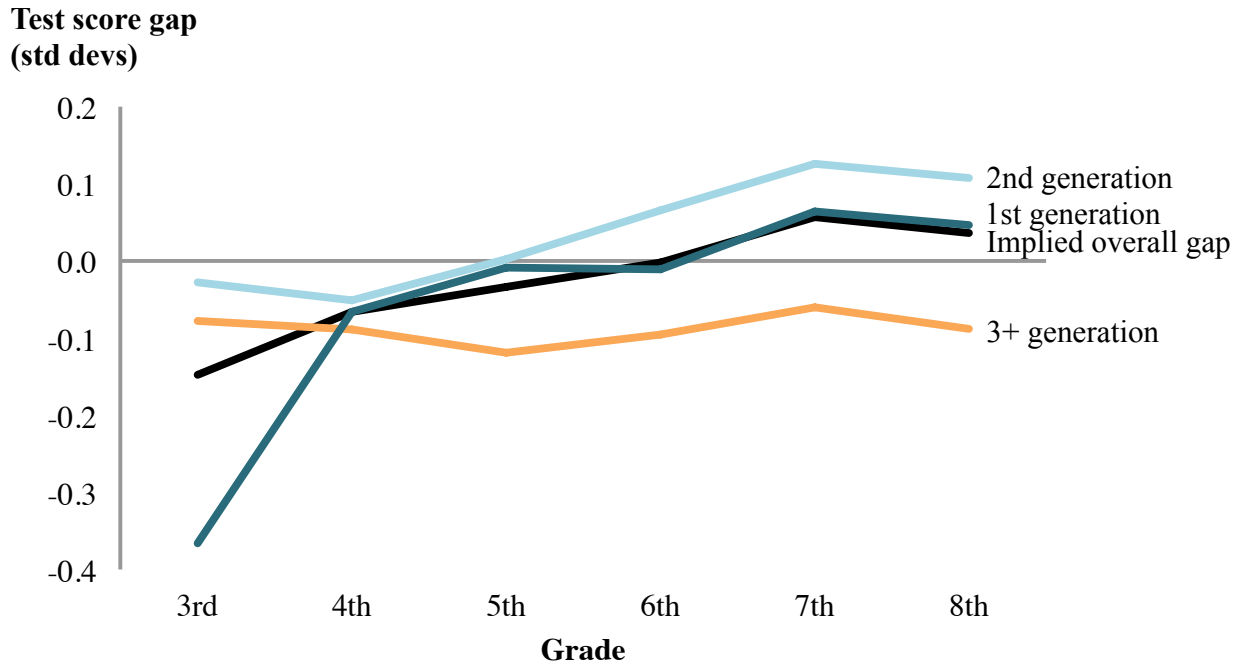
(B) Math



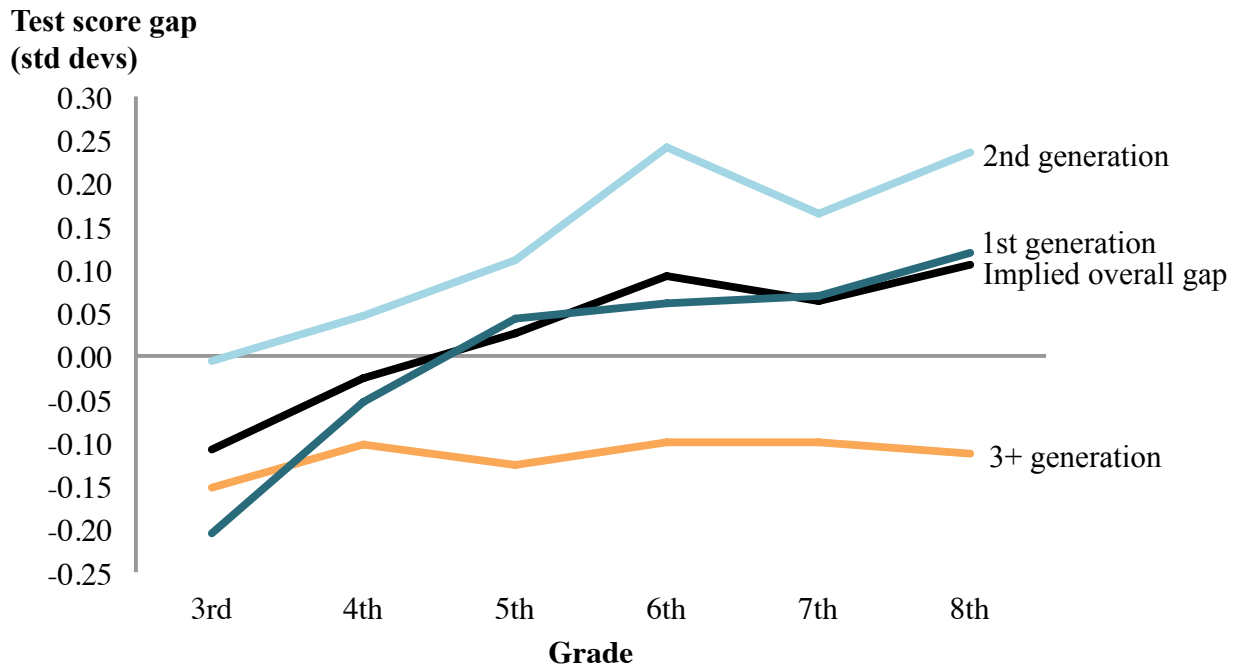
Sample: cohorts in 3rd grade in 2000 and 2001. Test score gaps are adjusted for late arrival and early exit, main effects and interactions with first-generation Hispanic. Regressions also include indicators for black, other race, multiracial, and missing other test score. See Tables 8 and 9 for point estimates and standard errors.

Figure 4: Adjusted Hispanic-white test score gaps by grade for continuous enrollees

(A) Reading



(B) Math



Sample: cohorts in 3rd grade in 2000 and 2001. Test score gaps are adjusted for parent's education (5 categories), free/reduced price lunch, gender, school-by-year fixed effects, and late arrival and early exit, main effects and interactions with first-generation Hispanic. Regressions also include indicators for black, other race, multiracial, and missing other test score. See Tables 10 and 11 for point estimates and standard errors.

A Weight adjustment appendix

This appendix explains in detail how and why I adjust the probability weights π_{ig} for under-matching between the birth data and education data. Of the Hispanic 3rd graders, 19% are matched to a birth record. Though this match rate is low, we would expect it to be fairly low given that first-generation Hispanics will not be matched by definition. Furthermore, some second- and third-generation students living in North Carolina were not born there. According to the Census/ACS, 37% of school-aged Hispanics living in North Carolina are first generation. Census/ACS data also reveal that only half of later-generation Hispanics were born in the state. Thus, I would expect to match 32% of the school records to a birth record, based on this second data source. In contrast, the actual match rate for the pooled sample of Hispanics in 3rd through 8th grade is 15%. (The 19% above refers to the proportion of 3rd graders matched. Since each year Hispanics enter NC public schools for the first time, this match rate declines for each grade.) While there are enough differences between the two samples (e.g. the exact age ranges considered) that I would not expect the the two rates to line up exactly, the actual and expected match rates are different enough to merit some concern.

There are two main implications of this data set matching problem. One, I do not know as many immigrant generations with certainty. Two, matching fewer second- and third-generation Hispanics means that my weights might be off. These later generations, specifically the ones that were born in North Carolina, essentially need more representation in the weights. A third concern is that the second- and third-generation Hispanics that are not matched to a birth record might be selected.

I adjust the probability weights to correct for this under-matching problem. With the correction, the weights reflect the fact that more second- and third-generation students are not matched to a birth certificate than is indicated by the Census/ACS. To make the adjustment, I perform the following calculation. First, drawing on what I expect the match rates by generation to be from the Census/ACS, I calculate the proportion of observations

that should have matched to a birth record but did not. I then add these fractions to the proportion of students by generation that I would not expect to match anyway. The relative proportions for each of these generations are the unconditional (on parent's education) weights that I want, i.e., the weights that account for the under-matching. For each generation, I divide this number by the proportion of that generation that was actually born out of state. Last, I multiply these weight adjusters by the sample weights provided by IPUMS (Ruggles et al., 2010). In my calculation of the probability weights by parent's education (i.e., the weights used in estimation), I use these adjusted sample weights.

To get a sense of whether the later-generation Hispanics that are not matched are selected, I compare the parent education distributions for the Hispanic students in the NCERDC that are known to be second- or third-generation with what those distributions should be based on the Census/ACS. This comparison is in Table A.1. I also show the corresponding distributions for students that are not matched. However the generations mix together for the unmatched Hispanics, the distributions of parent education in the NCERDC are not far off from the expected distributions. For the second- and third-generation students in the NCERDC that are matched, the biggest inconsistency is that parents with the lowest education level are overrepresented for the third generation. This check suggests that selection into birth record matching is not a concern.

Table A.1: Actual vs. expected parent education distributions

	NCERDC known 2nd gen	Implied from ACS ^a	NCERDC known 3rd gen	Implied from ACS ^b	NCERDC unknown gen	Implied from ACS ^c
Less than high school	0.584	0.619	0.249	0.158	0.470	0.432
High school graduate	0.350	0.312	0.518	0.486	0.411	0.359
Jr. college/trade school	0.038	0.016	0.117	0.161	0.054	0.084
Four-year college	0.024	0.040	0.099	0.135	0.054	0.084
Graduate school	0.005	0.014	0.018	0.061	0.011	0.052

NCERDC sample: pooled Hispanic students in grades 3-8 of the cohorts in 3rd grade in 2000 and 2001. Census/ACS sample: school-age Hispanic youth attending public school and living in North Carolina, 2000-2006. Used person-level weights provided by IPUMS.

^a Second-generation Hispanics born in North Carolina.

^b Third-generation Hispanics born in North Carolina.

^c Hispanics of any generation not born in North Carolina.

B Model fit appendix

In this appendix, I compare estimates of the overall Hispanic-white test score gaps implied by my model (the “implied” gaps) to those obtained with ordinary least squares (the “actual” gaps). I calculate the implied gaps from a weighted average of the Hispanic-white gaps by generation. By definition, the raw gaps in Table B.1 should be identical—the estimates come from the same sample and the models use the same controls—and indeed, they are very close. To compare the models that focus on continuous enrollees, I restrict the sample to students present in 3rd and 8th grade to estimate the actual gaps. Table B.2 documents that these estimates are very similar. Thus, my limited controls for late arrival and early exit seem to adequately account for movement in and out of the sample. Finally, I compare actual and implied overall gaps in Table B.3. These gaps line up well, with the exception of 7th and 8th grade reading. Still, the differences are less than 0.1 standard deviations in these two cases. These tables offer a check on my econometric model, with its parsimonious controls for continuous enrollment and iterative estimation of school-by-year fixed effects.

Table B.1: Actual vs. implied raw Hispanic-white test score gaps

Grade	3rd	4th	5th	6th	7th	8th
A. Reading						
Actual	-0.681	-0.682	-0.731	-0.760	-0.745	-0.813
Implied	-0.685	-0.685	-0.738	-0.766	-0.760	-0.832
B. Math						
Actual	-0.588	-0.558	-0.575	-0.608	-0.622	-0.606
Implied	-0.592	-0.562	-0.580	-0.613	-0.630	-0.614

Sample: cohorts in 3rd grade in 2000 and 2001. Actual gaps are estimated by OLS from a model with race/ethnicity and missing other test as controls. Implied gaps are calculated from a weighted average of the test score gap for each Hispanic generation from Tables 6 and 7.

Table B.2: Actual vs. implied raw Hispanic-white test score gaps for continuous enrollees

Grade	3rd	4th	5th	6th	7th	8th
A. Reading						
Actual	-0.686	-0.603	-0.577	-0.540	-0.480	-0.487
Implied	-0.679	-0.603	-0.590	-0.557	-0.500	-0.522
B. Math						
Actual	-0.593	-0.511	-0.465	-0.438	-0.454	-0.406
Implied	-0.588	-0.509	-0.474	-0.457	-0.476	-0.443

Sample: cohorts in 3rd grade in 2000 and 2001. Actual gaps sample only uses students present in 3rd and 8th grades. Actual gaps are estimated by OLS from a model that includes race/ethnicity and missing other test as controls. Implied gaps are calculated from a weighted average of the test score gap for each Hispanic generation from Tables 8 and 9.

Table B.3: Actual vs. implied adjusted Hispanic-white test score gaps for continuous enrollees

Grade	3rd	4th	5th	6th	7th	8th
A. Reading						
Actual	-0.159	-0.077	-0.047	-0.030	-0.034	-0.032
Implied	-0.148	-0.066	-0.033	0.002	0.056	0.036
B. Math						
Actual	-0.114	-0.034	0.013	0.073	0.057	0.115
Implied	-0.108	-0.025	0.026	0.092	0.064	0.106

Sample: cohorts in 3rd grade in 2000 and 2001. Actual gaps sample only uses students present in 3rd and 8th grades. Test score gaps are adjusted for parent's education (5 categories), free/reduced lunch, gender, and school-by-year fixed effects. Actual gaps are estimated by OLS from a model that also includes race/ethnicity and missing other test as controls. Implied gaps are calculated from a weighted average of the test score gap for each Hispanic generation from Tables 10 and 11.